User Request Evaluation Tool (URET) Conflict Prediction Accuracy Report

Mary Lee Cale
Michael Paglione
Dr. Hollis Ryan
Dominic Timoteo
Robert Oaks

April 1998

DOT/FAA/CT-TN98/8

U. S. DEPARTMENT OF TRANSPORTATION
Federal Aviation Administration
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

NOTICE

ERRATA

Report No. DOT/FAA/CT-TN98/8

USER REQUEST EVALUATION TOOL (URET) CONFLICT PREDICTION ACCURACY REPORT

April 1998

TECHNICAL NOTE

Prepared for
U. S. DEPARTMENT OF TRANSPORTATION
Federal Aviation Administration
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

REPLACED PAGES iii, 24, 26, 39, 48, 63

Released May 4, 1998

Acknowledgements

Efforts upon which the contents of this report are based were planned and conducted by the Traffic Flow Management branch (ACT-250) Conflict Probe Algorithm Assessment team. These team members include Mary Lee Cale and Dominic Timoteo, ACT-250; Michael Paglione and Dr. Hollis Ryan, Signal Corporation; and Robert Oaks, Crown Communications. Mrs. Cale was the FAA team lead for this effort. Mr. Paglione developed most of the metrics used in this report and was the principal architect and developer of the data reduction and analysis (DR&A) toolset. Dr. Ryan developed several of the DR&A tools and analyzed the conflict notification timeliness statistics and missed alerts in detail. Mr. Oaks documented the simulation process. Mr. Timoteo provided many thoughtful comments during the review of this document.

The overall effort required to successfully accomplish this task became very much an ACT-250 branch effort. The Conflict Probe Algorithm Assessment team members are very appreciative of the extra efforts extended by the following branch members:

Special appreciation goes to Bernie Day, Signal Corporation, who developed the Indianapolis simulation capability, without which this effort could not have been accomplished, and who cheerfully supported the team throughout many hours of simulation runs.

Tinh Vo, ACT-250, developed several software modules for the Conflict Analysis Tool (CAT) portion of the DR&A toolset.

Betty Kelly, CTA Incorporated, developed several software modules for the Track Conflict Probe (TCP) portion of the DR&A toolset and conducted much of the data analysis.

Jeff Jackson, Signal Corporation, developed the Oracle database system and supported its use during the simulation runs.

Edward Horvath, CTA Incorporated, was instrumental in the validation of the baseline simulation scenario and also supported many of the simulation runs and subsequent validation of results.

Joseph Pointkowski, CTA Incorporated, supported many of the simulation runs and the subsequent data analysis.

Finally, the support and cooperation of MITRE/CAASD in the use of their Algorithmic Evaluation Capability (AEC)/Xeval toolset was invaluable in accomplishing this effort. Xeval was used extensively to validate both the simulation scenarios and the analysis results. Special thanks go to Dan Brudnicki and Ruby Sharma.

Table of Contents

List of Figures and Tables

Executive Summary

This report presents quantitative measures for the conflict prediction accuracy of the User Request Evaluation Tool (URET) Delivery 3 (URET D3). These results are based on post-processing analysis of data collected during nine Indianapolis simulation runs of URET D3, in a single center operation, conducted by the Traffic Flow Management (TFM) Branch (ACT-250) at the FAA William J. Hughes Technical Center (WJHTC) in February 1998. Each simulation run had an average duration of five hours and contained between 450-500 simulated aircraft. The results provided here are consistent with those previously presented in the *URET Delivery 2.1 Conflict Prediction Accuracy Preliminary Report.*

The WJHTC simulation testbed emulates the URET field site installations (currently Indianapolis and Memphis Air Route Traffic Control Centers (ARTCCs)), with the National Airspace System (NAS) Host Computer System (HCS) connected to the URET system in the TFM laboratory via the fielded URET HCS interface, the General Purpose Output Interface Unit (GPOIU), but with no controllers in the loop. Use of a simulation allows ACT-250 to emulate what can not be observed or completely controlled in the real world, e.g., one would expect there to be no conflicts in the actual data since any potential conflict would have been resolved by the controller prior to its occurrence.

The metrics designed to assess the accuracy of URET's conflict prediction include conflict prediction accuracy and conflict notification timeliness. The underlying assumption of the metrics is that the HCS track data is the "ground truth". These metrics, presented in terms of probabilities to quantify the likelihood of correct and incorrect conflict predictions, include:

- Missed Alert – A conflict exists in the track data but URET did not present an alert to a controller.
- Strategic Missed Alert - Expands on the missed alert definition above to also include the case where a conflict exists in the track data and URET presented an alert to a controller less than five minutes before the start of the conflict.
- False Alert – A URET alert was presented to a controller but a conflict does not actually exist in the track data.
- Conflict warning time - Time difference between the time the URET alert is presented to the controller and the actual conflict start time (based on post-processing of track data).
- Conflict start time delta - Absolute value difference of the URET-predicted conflict start time and the actual conflict start time (based on post-processing of track data).

Results are presented for two analyses of the nine simulation runs, for a total of roughly 4500 aircraft per analysis. The first analysis, referred to as Analysis A, used standard separation as defined in FAA Order 7110.65 for en route airspace (i.e., five nautical miles in the horizontal dimension, and 1000 feet at or below FL290/2000 feet above FL290 in the vertical dimension). The second analysis, referred to as Analysis B, expanded the separation distance in the horizontal dimension to ten nautical miles which models the encounter distances URET uses in its predictions of yellow alerts for aircraft-to-aircraft conflicts. This method of analysis was specifically designed to allow the reader to decide on how these estimates should be used. The reader may choose values in either analysis, or gain insight into the sometimes subtle differences between the two. Furthermore, if the reader wishes to calculate other statistics based on the error events present in the simulations, two extensive appendices are supplied that contain all the partitioned individual counts from all nine simulations.

Reported missed alerts based on the two analyses had a very low probability (one missed alert was detected in Analysis A, and 13 were found in Analysis B). The grand average missed alert probability for all nine simulation runs is 0.002; for Analysis B it increased to 0.009. The strategic missed alert probability is higher at 0.04 for Analysis A, and 0.07 for Analysis B. The false alerts were much more common and possibly a reason for the low missed alert probability, since these two fundamental errors have an inversely proportional relationship (i.e., a false alert probability that is high by definition will provide a low missed alert probability). For Analysis A, the grand average false alert probability for both red and yellow alerts is 0.85. In other words, for the nine simulation runs processed for Analysis A, on average, the conditional probability that a given URET alert does not have a matching actual standard separation violation is 0.85. When the actual conflict definition was increased to ten nautical miles for Analysis B, this number was reduced by one fourth to 0.65. The false alert probability can also be conditioned by specific alert level (i.e., red or yellow). When the probabilities are conditioned in this manner for Analysis A, the conditional probability for a red alert being false is 0.72, and 0.93 for yellow. In other words, given a red alert, the probability that the alert is false is 0.72 for Analysis A. Similarly, for Analysis B, the conditional probability for a red alert being false is 0.51, and 0.73 for yellow. Finally, the grand average warning time is estimated at 15 minutes and the average conflict start time delta is 222 seconds (3.7 minutes) for Analysis A. There is a small reduction in these times for Analysis B (grand average warning time is estimated at 13.5 minutes and the average conflict start time delta is 130 seconds (2 minutes and 10 seconds), which is to be expected since an aircraft pair will be in violation of ten nautical mile separation earlier than for five nautical mile separation.

The technical accuracy of the URET conflict prediction algorithm is a critical issue to be considered in planning for additional field installations of the URET prototype as well as in preparation for a Joint Resources Council (JRC) investment decision for the Initial Conflict Probe (ICP). MITRE/CAASD has evaluated the accuracy and performance of the URET algorithms throughout the URET development effort and field evaluations, focusing on false alert statistics and predicted conflict warning time. This report provides quantitative statistics on URET missed alerts and *actual* conflict warning time, as well as on false alerts. This is valuable information for the decision makers charged with determining if the URET prototype should be installed in additional ARTCCs, as well as those making an investment decision for a production conflict probe. It should also prove useful to both the developer of the URET prototype and the conflict probe production contractor. It is recommended that additional studies be conducted, using actual field data adjusted to include predefined conflict situations, to expand the conflict accuracy estimation to include measurements of trajectory accuracy and conflict prediction stability.

1. Introduction

This report presents quantitative measures for the accuracy of the User Request Evaluation Tool (URET) conflict prediction, for aircraft-to-aircraft conflicts in Current Plans, in terms of the following metrics: missed alerts, false alerts, and conflict notification timeliness. These results are based on analyses of data collected during nine Indianapolis simulations of URET Delivery 3 (D3) in single center operation conducted by the Traffic Flow Management Branch (ACT-250) at the FAA William J. Hughes Technical Center (WJHTC).

1.1 Concept/Background

URET is an automated conflict detection (ACD) tool that is currently in use in prototype form as a decision support aid for the en route radar associate ("D-side") air traffic controllers at the Indianapolis (ZID) and Memphis (ZME) Air Route Traffic Control Centers (ARTCCs). URET detects aircraft-to-aircraft and aircraft-to-airspace conflicts for IFR aircraft tracked by the Host Computer System (HCS), and provides alert information to the D-side controller when such conflicts are detected. The URET prototype is considered to be a risk reduction activity for the Initial Conflict Probe (ICP) on the Display System Replacement (DSR).

The technical accuracy of the URET conflict prediction algorithm is a critical issue to be addressed in planning for possible additional field installations of the URET prototype as well as in preparation for a Joint Resources Council (JRC) investment decision for ICP. The MITRE/Center for Advanced Aviation System Development (CAASD) has evaluated the accuracy and performance of the URET algorithms throughout the URET development effort and field evaluations, focusing on false alert statistics and predicted conflict warning time. For future URET/ICP decisions, the FAA requires quantitative statistics on URET missed alerts and **actual** conflict warning time as well as false alerts.

1.2 Scope

As part of previous activities, ACT-250 developed a detailed set of generic metrics to be used to evaluate the accuracy of existing conflict prediction tools (References: *Generic Metrics and Statistics to Estimate the Conflict Prediction Accuracy of Conflict Probe Tools* and *Plan for Evaluation of the Conflict Probe Programs*). These metrics, which include trajectory accuracy, conflict prediction accuracy, conflict prediction stability and conflict notification timeliness, were designed to be applied to any conflict prediction tool available, thereby providing common measures to evaluate the performance of different systems. The results presented in this report focus on a subset of these generic metrics, i.e., conflict prediction accuracy and conflict notification timeliness (see Section 2.2), applied to URET D3's prediction of aircraft-to-aircraft conflicts in Current Plans for a single center operation.

This page intentionally blank.

2. Overview

ACT-250's approach in accomplishing this effort was to develop a ZID simulation capability at the WJHTC. This approach was chosen, rather than using actual "real world" data, because one would expect there to be no conflicts in the actual data since any potential conflict would have been resolved by the controller prior to its occurrence. Use of a simulation allows the emulation of situations that can not be observed or completely controlled in the real world, and allows a wide range of test cases to be generated and analyzed by ACT-250 using the basic infrastructure and analysis tools. Another advantage of this approach in assessing the accuracy of conflict prediction tools is that the conflicts can be modeled at any minimum separation desired.

The WJHTC simulation testbed emulates the URET field installations with the NAS Host Computer System (HCS) connected to the URET system in the Traffic Flow Management (TFM) laboratory via the fielded URET HCS interface, the General Purpose Output Interface Unit (GPOIU). The simulation process begins with a System Analysis Recording (SAR) from a field facility, in this case ZID. The flight plans extracted from the SAR are input to the Pseudo Aircraft System (PAS) which simulates the aircraft flight path and generates simulated aircraft position data. This position data is then transmitted to the WJHTC Target Generation Facility (TGF) which creates simulated radar reports. The WJHTC HCS receives the TGF radar messages, as well as flight plan information extracted from the ZID SAR, and provides flight plan and simulated track data to URET in the TFM laboratory via the GPOIU. During the simulation run process, the HCS track reports are recorded by the Monitor Test and Recording (MTR) module attached to the GPOIU. This MTR file, which contains all the track messages received by URET, provides the "ground truth" track reports. Using URET's data recording capability, alert records are also recorded. These two source files (i.e., what actually happened from the tracks and what was predicted from the alerts), are utilized during post-processing by the data reduction and analysis tool set (discussed in Section 2.3) to estimate the conflict prediction accuracy of URET. An overview of this complete process is depicted in Figure 2.0-1.

ACT-250 is developing the simulation capability in an incremental manner. The current capability, referred to as Phase One, was designed to provide quantifiable data to support an ICP JRC decision. As such, this initial effort focused on an experiment designed to measure URET's conflict prediction accuracy, specifically the probability of missed alerts and false alerts, as functions of actual aircraft horizontal and vertical separations, and conflict notification warning time. During Phase Two, ACT-250 plans to estimate the complete set of generic metrics (identified in Section 1.2) to assess the conflict prediction accuracy of URET, as well as other conflict prediction tools. Multiple simulations will be run under varying conditions and field data will also be analyzed. Sensitivity testing will also be conducted. Future efforts will include more detailed designed experiments to answer specific questions on the effects of various factors (e.g., encounter geometry, flight path characteristics, etc.) on the metrics.

2.1 Simulation

A URET simulation testbed has been established in the TFM laboratory at the FAA WJHTC. Development of the simulation capability to collect the required data for analysis is composed of two processes: generation of the simulation scenario data and the actual simulation runs during which the data is collected. These processes are discussed in the following sections.

2.1.1 Scenario Generation Process

As depicted in Figure 2.1-1, the scenario generation process began with SAR tapes recorded at ZID in July 1997. These SAR tapes, which contained approximately six hours of data, were processed by a SAR Extraction process developed previously at the WJHTC. This process extracted the flight plans for aircraft arriving at or departing from airports located in, and overflights traversing, the seven ZID sectors which Air Traffic had identified to be of interest: Pocket City, Louisville, Rebel and Falmouth high altitude sectors, and Evansville, Nabb, and Hazard low altitude sectors. As a result, all extracted aircraft flew in adherence to their original filed flight plans without redirection due to controller actions (i.e., any subsequent controller actions contained in the SAR data that may

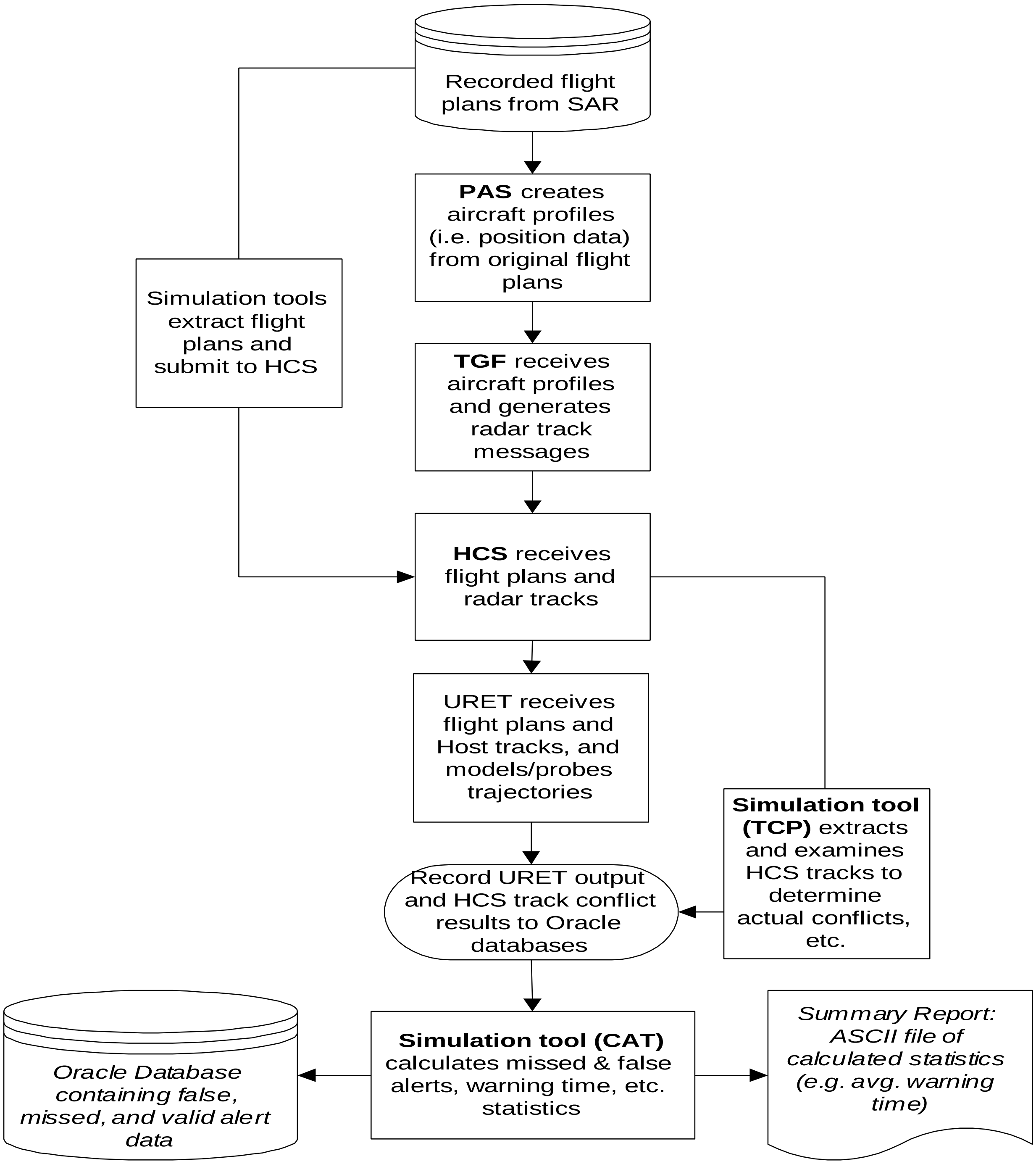


Figure 2.0-1: Simulation Overview

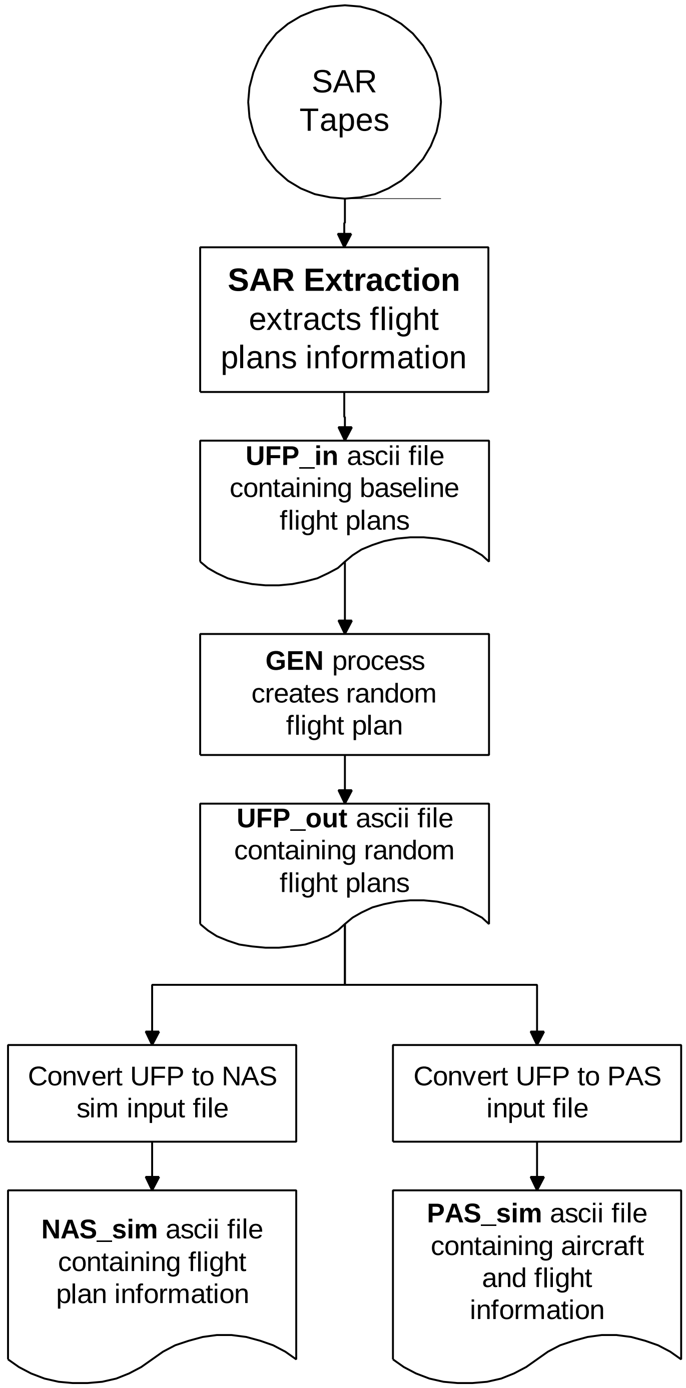


Figure 2.1-1: Scenario Generation Process

have affected the aircraft's flight were not included). In addition, some specific modifications were made to the standard SAR Extraction process for this study:

- Coordination fixes outside the sectors were used as start points rather than the x-y coordinates at the handoff location.
- Departures climbed directly to their flight plan altitudes rather than to an intermediate assigned departure altitude.
- Arrivals were started at assigned altitudes and then descended to the arrival airport.

The SAR Extraction process created an initial *UFP_in* file formatted in the Universal Flight Plan (UFP) format. Validation of the scenario contained in the initial *UFP_in* file was a time-consuming process that was facilitated by the use of the MITRE/CAASD-developed Algorithm Evaluation Capability (AEC)/Xeval tool set. During the scenario validation process, additional aircraft were extracted because of discrepancies between the PAS and URET adaptations. The baseline scenario that resulted was about four hours in length and contained 257 aircraft. The conflict situations for the 257 aircraft were determined to be too few for meaningful analysis. Therefore, an existing in-house simulation tool (the GEN program, also previously developed at the WJHTC) was used to create additional realistic scenarios from the baseline scenario. This process, discussed in the following paragraph, produced a scenario with a heavier traffic load, more accurately reflecting a typical load found in ZID today. This resulted in a significant increase in the number of conflict situations (Note: Using this approach, the simulation produces more conflicts and thus alerts at a higher rate than the real system).

The GEN program was used to generate a UFP formatted file (identified as *UFP_out* in Figure 2.1-1) for each simulation run. The GEN program accepted the *UFP_in* file as input, along with desired scenario parameters such as the duration of the simulation and the number of aircraft to be simultaneously active (for this study these were five hours and 90 aircraft). The GEN program created each *UFP_out* file by first introducing aircraft into the simulation at a constant rate during a ramp-up time (15 minutes). This resulted in the introduction of one aircraft every 10 seconds (15 minutes/90 aircraft). Once the desired quantity of aircraft was reached another aircraft was not introduced until an aircraft left the simulation, either by landing or flying beyond the center boundary. The algorithm which identified the aircraft to add to the simulation randomly drew aircraft from the baseline flight plans so that:

- The percentage of aircraft in the simulation closely matched the distribution of flight plan categories found in the baseline sample (categories represented selective categorization of the flight plans by similar characteristics; e. g.; arrivals, departures, routes, etc.).
- All of the baseline flight plans were used at least once during a simulation run.

The simulation runs required the conversion of these *UFP_out* files to two different files: one for PAS and the other for the HCS. These files, identified in Figures 2.1-1 and 2.1-2 as *PAS_sim* and *NAS_sim*, were the input to the simulation run process discussed in the following section.

2.1.2 Simulation Run Process

As previously mentioned, the simulation run process (depicted in Figure 2-1.2) involved the URET system located in the TFM laboratory, as well as three other established systems located at the WJHTC: the PAS, the TGF, and the HCS (NAS version 1.3). These systems were adapted for ZID based on the May 22, 1997 ZID Adaptation Controlled Environment Subsystem (ACES) output. The ZID standard operating procedures (SOPs) and letters of agreement (LOAs) were also utilized in the adaptation process. All procedural altitude restrictions were turned off in URET during the runs, and the simulation did not model radar noise or wind.

PAS simulated the aircraft flight paths specified in the *PAS_sim* input file and generated simulated aircraft position data which was transmitted to the TGF. TGF created simulated radar reports which were transmitted to the HCS. The HCS received the TGF radar messages, as well as flight plan information contained in the non-radar simulation tape created from the *NAS_sim* file, and provided flight plan and track data to URET via the GPOIU. Each of these simulation runs lasted about five hours, during which data was recorded from both the HCS and URET. The HCS track reports were recorded by the MTR module attached to the GPOIU. This MTR file, which contains all the track messages received by URET, provided the "ground truth" track reports. Using URET's data recording capability, URET alert records were also recorded. These files were the input for the data reduction and analysis phase of the study (see Section 2.3).

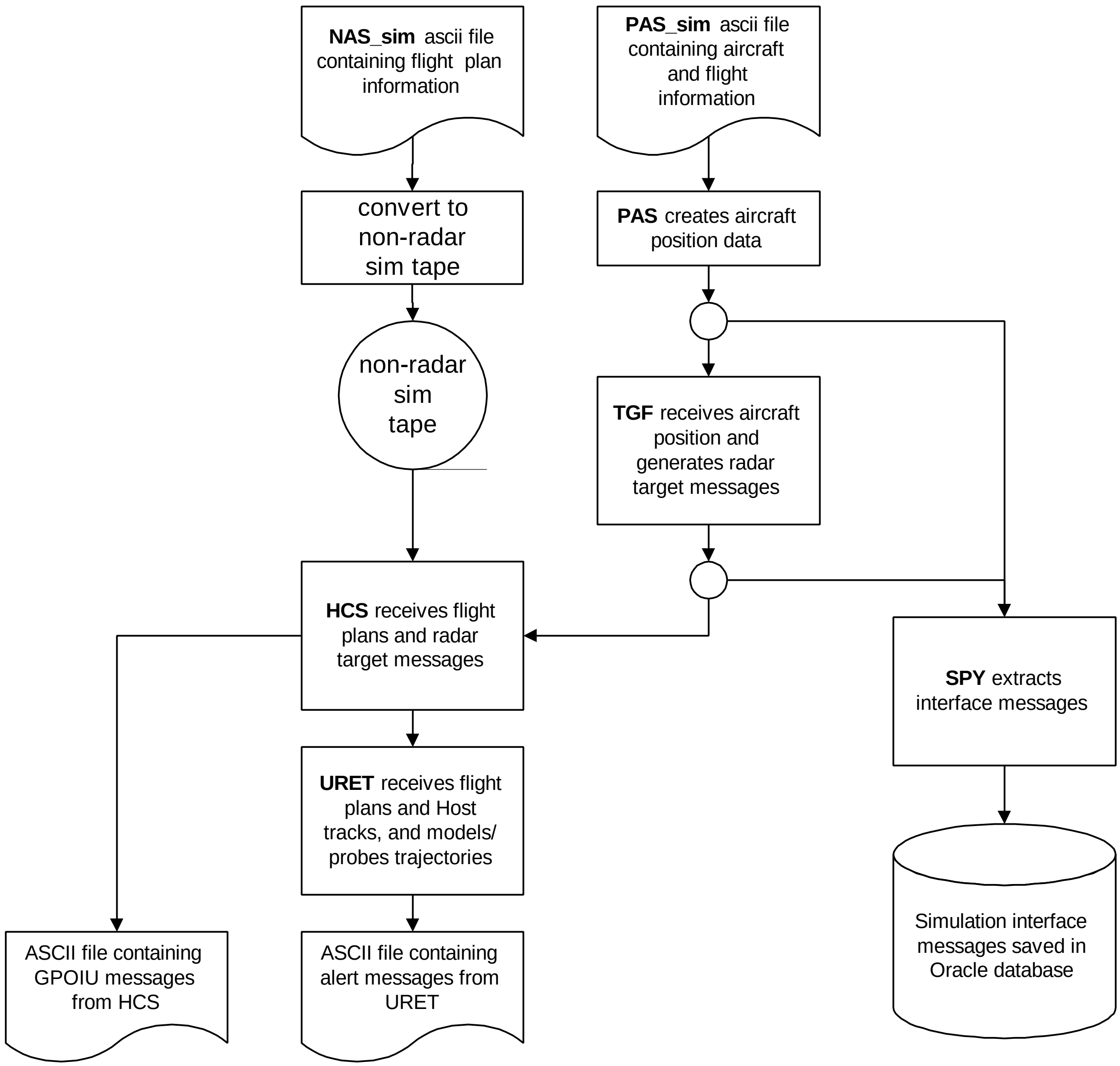


Figure 2.1-2: Simulation Run Process

Additional simulation interface data was recorded during the run using the SPY process developed within ACT-250 to support the TFM Laboratory. This process extracts data transmitted between the PAS and TGF interface, and the TGF and HCS interface, and places it in the TFM Oracle database. While this data was not used directly in the study, it was available whenever specific issues arose.

2.2 Metrics

The metrics designed to assess the accuracy of URET's conflict prediction include conflict prediction accuracy and conflict notification timeliness. These metrics, presented in terms of probabilities, will quantify the likelihood of correct and incorrect conflict predictions. The underlying assumption of the metrics is that the HCS track data is the "ground truth". Details on the exact equations used to determine the accuracy statistics are found in *Generic Metrics and Statistics to Estimate the Conflict Prediction Accuracy of Conflict Probe Tools.*

2.2.1 Conflict Prediction Accuracy

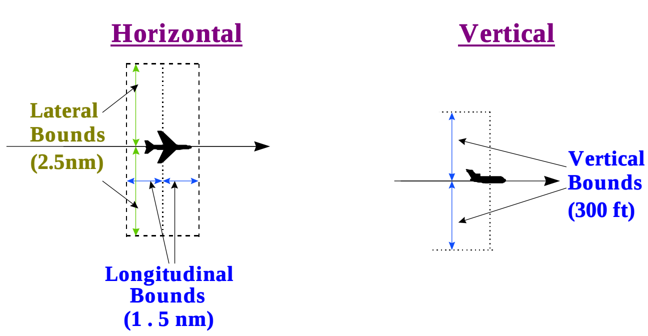
The accuracy of the conflict prediction is basically the measure of the difference between what URET predicts will occur (based on its' predicted trajectory) and presents to the D-side controller as a URET alert (red or yellow), and what is determined to have actually occurred from post-processing of the HCS track messages. URET presents two levels of alerts to the D-side controller for predicted aircraft-to-aircraft conflicts: red alerts indicate that the centerlines of two URET aircraft trajectories are predicted to be in violation of standard separation, while yellow alerts indicate that the conformance boxes[1] that surround the aircraft's trajectories are predicted to be in violation of separation standards. For purposes of this report, there is no differentiation between regular red or yellow alerts and "muted" red or yellow alerts.

The conflict prediction accuracy metrics are designed to describe two fundamental random events: a conflict exists that is not predicted in a timely manner or at all, or an alert is presented when there is not really a conflict. These events, whose outcomes are summarized in Table 2.2-1, require the following definitions:

- Conflict – A violation of separation standards between two aircraft was determined from the HCS track data during simulation post-processing.
- Valid Alert – An alert (red or yellow) was presented to a controller by URET and a conflict exists in the track data (Note: an alert is considered to occur only if URET presented a predicted conflict to a controller and it is recorded in the DCR_PDM_CP_ALERT record by URET's data recording software).
- Missed Alert – A conflict exists in the track data but URET did not present an alert to a controller.
- Late Valid Alert - A conflict exists in the track data and URET presented an alert to the controller less than five minutes before the start of the conflict (at least five minutes of track data must exist for this to occur).
- False Alert – An alert (either red or yellow) was presented to a controller by URET but a conflict does not actually exist in the track data.

---

[1] URET places a conformance box around each aircraft, centered on the aircraft trajectory, to represent regions of uncertainty. The conformance box has the following dimensions for aircraft in straight and level flight (depicted below): 2.5 nautical miles lateral, 1.5 nautical miles longitudinal, and 300 feet vertical. The conformance box is expanded in the appropriate dimension(s) when an aircraft is turning or climbing/descending, and for non-RNAV equipped aircraft.



	CONFLICT OCCURS	CONFLICT DOES NOT OCCUR
ALERT	URET predicts conflict/ conflict occurs (Valid Alert/Late Valid Alert)	URET predicts conflict/ conflict does not occur (False Alert)
NO ALERT	URET does not predict conflict/ conflict occurs (Missed Alert)	URET does not predict conflict/ conflict does not occur

Table 2.2-1: URET Alert/Actual Conflict Events

For these results, the concept of a strategic missed alert is also introduced. A strategic missed alert (SMA) expands on the missed alert calculation to include late valid alerts (Note: SMA is similar to the ICP definition of missed alert as defined in Section 3.2.1.2.5 of the *ICP/Enhanced Display System Infrastructure (EDI) System/Segment Specification, Volume I, Part 2*). Metrics on missed alert events are represented by two statistics in this report: true Missed Alerts and Strategic Missed Alerts. Estimates of the strategic missed alerts would be expected to be equal to or higher than the original missed alert probability, since SMA includes missed alerts based not just on their presence but on their timeliness of being presented.

The fundamental error events, missed alert, strategic missed alert and false alert, are modeled by probability metrics that quantify the chances or likelihood of these events occurring. First, the probability of a missed alert (MA) is defined by the conditional probability that an aircraft conflict will not be presented to a controller (i.e., produce an alert) at all by URET. That is, the probability of MA is the probability that no alert is presented given an actual conflict occurs. The probability of a strategic missed alert (SMA) is defined by the conditional probability that an aircraft conflict will not be presented to a controller (i.e., produce an alert) at all by URET or will be presented too late for strategic use; i.e., the probability of a SMA is the combined probability that, given the occurrence of an actual conflict, no alert is presented to the controller or a valid alert is presented with less than five minutes warning time. Finally, the probability of a false alert (FA) is defined as the conditional probability that an aircraft alert will have no conflict. In other words, the probability of FA is the probability that no conflict exists given URET presents an alert.

2.2.2 Conflict Notification Timeliness

Conflict notification timeliness is measured in terms of the actual conflict warning time. The actual conflict warning time for each valid alert is the time difference in seconds between the time the URET alert is presented to the controller (i.e. notified of an alert) and the actual time when the two aircraft are first determined to be in conflict (based on the track data).

As implied in the definition, conflict warning time is calculated only for valid alerts (not the subset of valid alerts called late valid alerts). There are valid alerts with corresponding conflicts in our analysis that occur very early in recorded tracking of the flight (e.g., pop up situations). These valid alerts are excluded from the calculation of the warning time statistics since URET would not be able to provide sufficient warning time for the impending conflict situation. The current rule used for excluding such valid alerts consists of determining if the conflict start time for these alerts is less than a parameter time (e.g., five minutes) from the start of the track data.

Related to the conflict notification timeliness is the conflict start time delta statistic. Conflict start time delta is the absolute value difference of the predicted start time of the conflict and the actual conflict start time. Thus it is a measure of the prediction error for the start of the conflict. Like conflict warning time, conflict start time delta is calculated only for valid alerts that have sufficient track data from the start of the conflict (i.e., five minutes of track data) and does not include the subset of valid alerts called late valid alerts.

2.3 Data Reduction and Analysis

The effective estimation of the metrics discussed in Section 2.2 requires considerable data to be collected and analyzed. This data collection process was automated by the development of an Oracle database system and an extensive generic data reduction and analysis (DR&A) tool set. The DR&A tools determine the difference between the actual track conflicts and the URET predicted conflicts, and compute the probability metrics; i.e., the accuracy of URET's conflict prediction. This tool set (delineated in Figure 2.3-1) is comprised of the Track Conflict Probe and the Conflict Analysis Tool, described in the following sections. Due to the complexity of the flight profiles, the MITRE/CAASD-developed AEC/Xeval graphical program was also used extensively to validate many of the error events (i.e., missed alerts and late valid alerts) and simulated flight profiles.

2.3.1 Track Conflict Probe (TCP)

The Track Conflict Probe (TCP) determines the actual conflicts that are in the scenario based on the "ground truth" HCS track data contained in the Oracle database. TCP first takes the raw track data, which is in 12 second increments, and time synchronizes it into parameter size increments (e.g., ten seconds) using a quadratic interpolation function (refer to ACT-250 *URET Algorithm Assessment Report*, Section 3.1.9). This technique allows the analyst to interpolate through missing track points or to adjust the time to different size increments. The time synchronized track reports are then stored in two Oracle databases, one which contains the actual time synchronized track reports and one which contains summary information for the track reports (e.g., maximum and minimum distances per dimension, etc.).

Once the track reports have been time synchronized, each track data point is examined to determine if it lies within an adapted automated problem detection inhibited area (APDIA). The APDIA is a volume of airspace, usually surrounding terminal areas, in which URET does not probe for conflicts. TCP uses an algorithm similar to that used by URET (GM_REGN) to make this determination, and prohibits conflict detection and minimum separation calculations for track position reports determined to be in an APDIA.

For those aircraft whose track data is not found to be within an APDIA, TCP then looks at all n(n-1)/2 aircraft pairs to determine if there is some time overlap in the track data. Those aircraft with some time overlap are then subject to a gross filter which utilizes the track summary information contained in the Oracle database to determine whether horizontal and vertical separation requirements have been violated. Basically, this is accomplished by constructing three-dimensional rectangles around the entire track set for each aircraft and computing the separation between the rectangles. Aircraft pairs whose rectangles are separated by less than 25 nautical miles in the horizontal dimension, and less than 5000 feet in the vertical dimension, are considered to "pass" the gross filter.

Only those aircraft pairs that pass the gross filter are checked for minimum separation violations based on the specified separation standard. Separation violations are determined using an iterative process of comparing track coordinates at equivalent times and calculating a simple set of ratios. These ratios consist of calculating the horizontal separation divided by standard horizontal separation for the particular point in space, and the vertical separation divided by standard vertical separation. The maximum of these two ratios is referred to as the max-ratio. If the max-ratio is less than one, a violation of separation must have taken place. The TCP tool iteratively determines the minimum max-ratio from all time overlapping track points for each aircraft pair that passed the gross filter. If this minimum max-ratio statistic is less than one, a conflict is determined to have occurred (The minimum max-ratio statistic will be further applied in future phase two studies for sensitivity analysis; e.g., sharpness metric as defined in *Generic Metrics and Statistics to Estimate the Conflict Prediction Accuracy of Conflict Probe Tools*. Preliminary information on the application of the sharpness metric to the current analysis of URET D3 is provided in Appendix C).

Two Oracle databases are produced at the completion of TCP: aircraft pairs in conflict and aircraft without conflicts. For both of these databases, two basic sets of separation distances and associated times are calculated for each pair of aircraft. The minimum horizontal separation (MHS) is the minimum distance in the horizontal dimension, measured in units of nautical miles. The minimum vertical separation (MVS) is the minimum distance in the vertical dimension, measured in units of feet. The MHS and MVS can occur anywhere along the track data sets of the aircraft; i.e., the MHS and MVS can occur at completely different times, since each dimension, either vertical or horizontal, represent separate processes. For aircraft pairs in conflict, the MHS and MVS are calculated for the duration of the conflict.

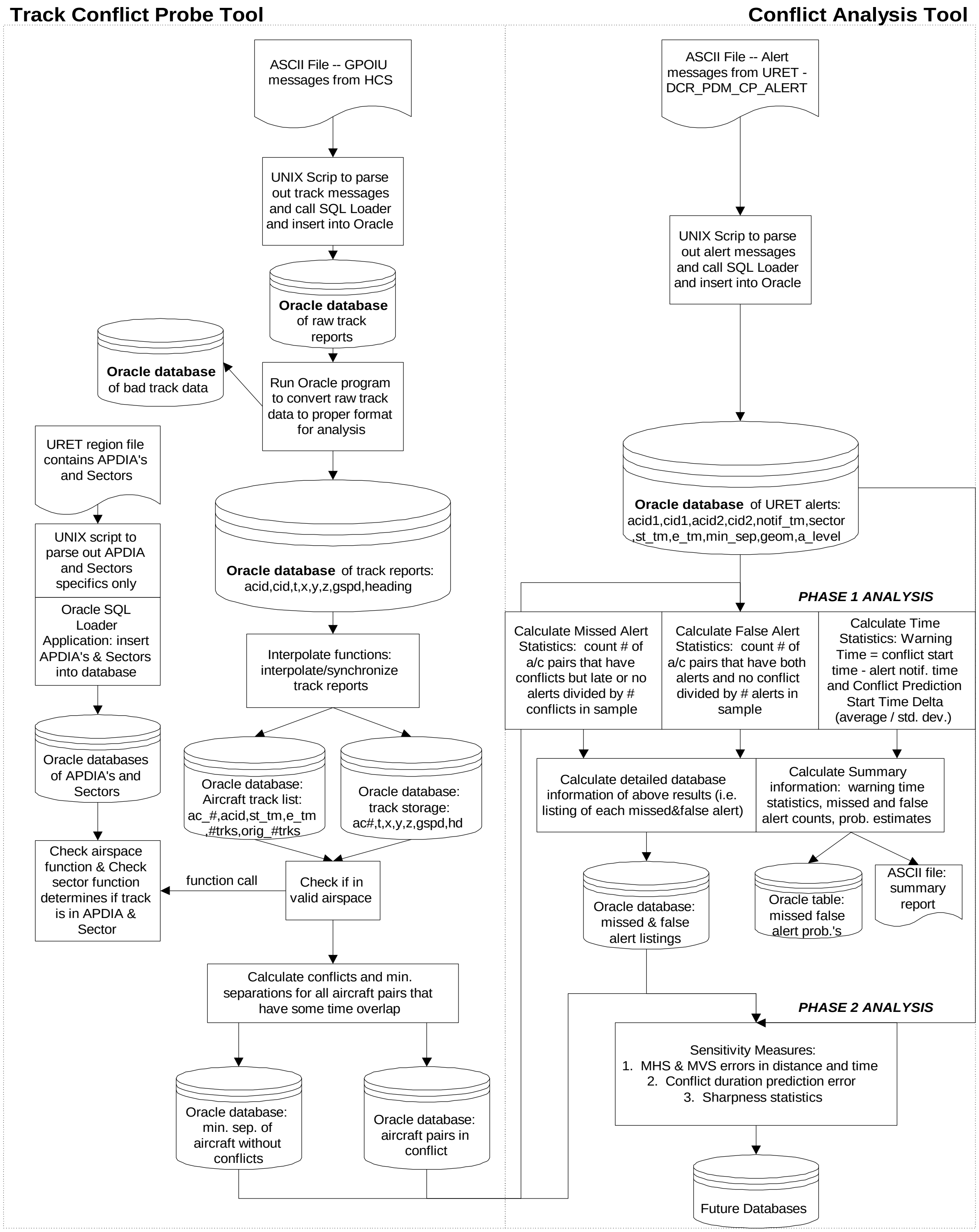


Figure 2.3-1: Overview of DR&A Tool Set

TCP currently employs two rules to exclude aircraft pairs from being considered in conflict:

- The duration of the conflict is less than a parameter threshold (e.g., ten seconds, in which case it is based on only one track report).
- A pair of aircraft in the cruise phase of flight has a minimum vertical separation up to but not including 300 feet less than standard vertical separation. (This exception attempts to model the case where two aircraft are cleared for cruise at different flight levels providing standard separation, however one aircraft is tracked less than 300 feet off the cleared altitude. This could potentially be a violation of standard separation, but unless the flight is more than 300 feet off the cleared altitude NAS does not consider them in conflict.)

2.3.2 Conflict Analysis Tool (CAT)

The Conflict Analysis Tool (CAT) compares the output of TCP with the alerts notified to a controller by URET (contained in the Oracle database, extracted from the URET DCR_PDM_CP_ALERT "add" records), and computes the following statistics:

- Probability of missed alert
- Probability of strategic missed alert
- Probability of false alert
- Conflict warning time
- Conflict start time delta

Basically, CAT examines each conflict determined by TCP and looks for a corresponding URET alert in the Oracle databases. Only unique, notified URET alerts are counted as alerts (i.e., new alerts as opposed to updates to existing alerts). Those conflicts that have no corresponding URET alert are considered Missed Alerts. Alerts which have a corresponding conflict identified are flagged as Valid Alerts; duplicate alerts are excluded from the analysis. The subset of Valid Alerts determined to have conflict warning times less than five minutes are considered Late Valid Alerts. By process of elimination, all remaining alerts for which there is no corresponding conflict are initially considered to be False Alerts. The false alerts are then filtered for exceptions based on the following set of rules:

- URET alerts that predict a conflict outside the available track data for either involved aircraft are eliminated (i.e., the URET predicted conflict begins or ends before/after the existence of "ground truth" track data).
- URET alerts for aircraft that never had any track data (and therefore no "ground truth" data for TCP to use to determine a conflict) are excluded from analysis. These occurrences are very rare and are an artifact of the simulation process; about five aircraft were never tracked by the TGF/HCS.
- URET alerts posted beyond the end of the scenario run, or the last track report time, are eliminated. This is another artifact of the simulation process, since URET will continue presenting alerts based on flight plan data even after the HCS stops sending messages marking the end of the simulation. By eliminating alerts past the last track report, the end of all simulations is standardized and protected against this artifact.

The process of matching an alert with a conflict can be more complicated if there is more than one notified alert for a particular aircraft pair in conflict. In other words, the CAT requires matching logic to determine which alert to consider the Valid Alert when more than one alert exists. At the time the CAT was initially developed, the URET alert records contained only time prediction data not the physical coordinates of the conflict. The matching rule used for this study matches the first notified alert for corresponding aircraft callsigns to the specific conflict. This insures that the notification, and thus the warning time, is optimal for the particular conflict situation. Later, an additional set of matching logic for Valid Alerts was developed for the CAT program. This second rule, which may be used in future studies, chooses the alert with the minimum conflict start time deviation or delta (i.e., absolute difference between conflict start prediction and actual conflict start times) with corresponding aircraft callsigns. Therefore, the second rule, while not providing the best warning time, will provide the minimum conflict start time delta.

The CAT function computes the following statistics:

- probability of Missed Alert = number of Missed Alerts/number of conflicts
  (note that the number of conflicts is equal to the sum of Missed Alert and Valid Alert counts, including Late Valid Alerts)
- probability of Strategic Missed Alert = number of Strategic Missed Alerts/number of conflicts
  (note that the number of Strategic Missed Alerts is equal to the sum of Missed Alert and Late Valid Alert counts)
- probability of False Alert = number of False Alerts/number of unique alerts
  (note that the number of unique alerts is equal to the sum of Valid Alert and False Alert counts)
- Conflict Warning Time = actual conflict start time **-** URET alert notification time
- Conflict Start Time Delta = absolute value of the difference between actual conflict start time and URET-predicted conflict start time; i.e., | actual conflict start time **-** predicted conflict start time |

2.3.3 Limitations of the Data Reduction and Analysis Tools

The data reduction and analysis software, which was developed and used to automatically process the simulation data, is limited in its ability to correctly process all of the data. As a result, some manual verification and data editing is necessary to supplement the automatic processing. One example of this manual process is the situation where two aircraft cleared at different flight levels and crossing horizontally have a minimum vertical separation slightly below 2000 feet (but greater than 1700 feet). If the lower aircraft starts a descent close to the end of the encounter, this aircraft-to-aircraft encounter is currently evaluated as a conflict by the DR&A tools when there would not really be a conflict from a controller's point of view. Since the aircraft pair were not only cleared at vertical separation but diverging, they were manually excluded as conflicts and the associated incorrectly assessed missed alerts were also removed. In another example, both aircraft in a nominal conflict are just grazing an APDIA, which may be caused by track and interpolation inaccuracies. URET would not consider this a conflict if they were considered to be in the APDIA, in which conflict probing is inhibited. In the future, more software logic will be added to the DR&A tools to address these rare anomalies.

This page intentionally blank.

3. Results and Observations

The results presented in this section are based on post-processing of nine ZID simulation runs conducted at the WHJTC in February 1998. Each run had an average duration of five hours and contained approximately 450-500 simulated aircraft. The total number of aircraft for all the simulation runs was thus around 4500, yielding over a million aircraft-to-aircraft combinations. All procedural altitude restrictions were turned off in URET during these runs, and the simulation did not model radar noise or wind.

The nine simulations were analyzed twice with two different definitions of aircraft-to-aircraft conflicts. The first analysis, referred to as Analysis A (see Section 3.1), used basic standard radar separation for en route airspace as defined in FAA Order 7110.65, 4-5-1.a/b and 5-5-3.b.1 (i.e., five nautical miles in the horizontal dimension, and 1000 feet at or below FL290/2000 feet above FL290 in the vertical dimension). This represents the true standard separated conflict situation. The second analysis, referred to as Analysis B (see section 3.2), expands the separation distance in the horizontal dimension to ten nautical miles, which more closely models the encounter distances URET uses in its predictions of yellow alerts for aircraft-to-aircraft conflicts.

The two conflict definitions (A and B) have expected effects on the various statistics, which are discussed in more detail in Sections 3.1 and 3.2. In general, it was expected that more conflicts would occur, and occur earlier, if the conflicts were considered at larger separation distances. This is exactly what happened in this study. By providing two analyses, the reader can decide which analysis is best suited for the particular statistic. For example, it is more critical to examine the missed alert probability of Analysis A, since this represents missed conflict predictions for actual violations of standard legal separations. However, for the false alert statistics, Analysis B represents a better view of the performance of a conflict probe tool designed to present alerts at more than standard separation.

Grand averages and confidence intervals for all nine simulations were computed for the statistics discussed in section 2.2. The grand average, which provides a point statistic for each metric, is the average of all the averages for each statistic for the nine simulation runs (e.g., average warning time was calculated for each simulation run and then a grand average for the warning time for all the simulations was calculated). A confidence interval with a 90 percent confidence level was also calculated for each statistic. This confidence interval provides the upper and lower limits of the statistic with 90 percent confidence (i.e., the probability of including the true value of the statistic). The confidence interval weights the variation of the point statistic, as well as the central tendency, and provides a probability measure of the desired confidence. For example, if the warning time changed significantly from one simulation to the next, the confidence interval would be wider than if it did not change much at all.

Sections 3.1 and 3.2 first present the scenario summary information totals (e.g., the total number of aircraft in all the simulations, the average number of aircraft, etc.), the error probability tables, the notification timeliness statistics (e.g., warning time), and the conflict prediction start time deviations. The probability tables are presented as averages, and then separate tables are provided for the upper and lower confidence interval bounds. Both the average and confidence intervals for the time statistics (i.e., warning time and conflict prediction start times) are presented in one table. A discussion of specific observations as to why certain error events (i.e., missed alert, late valid alert, false alert) may have occurred is provided in section 3.3. The specific counts for each simulation and grand average counts for all the runs for analysis A and analysis B are provided in Appendices A and B, respectively.

3.1 Analysis A

The results of the data reduction and analysis based on standard separation distances between track reports are presented in the following tables. Tables 3.1-1 and Table 3.1-2 present the scenario information and various alert events (i.e., missed alert, false alert, strategic missed alert) for the nine simulation runs in terms of grand totals and grand average values. Of the 4412 total aircraft, there was an average of 490 aircraft per simulation run. The number of aircraft pair combinations is equal to [n(n-1)]/2 where n is the number of aircraft per simulation. The average number of aircraft combinations per simulation is thus around 120,000. Of these 120,000 aircraft pairs, only around 31,000 on average had some type of time overlap. Of the time overlapping aircraft pairs, about 18,500 pairs passed the TCP conflict gross filter. As described in Section 2.3.1, the gross filter constructs rectangles around all the track points of a given aircraft and then compares these rectangles. If they are within 25 nautical miles in the horizontal and 5000 feet in the vertical, they *pass* the gross filter. Therefore for Analysis A, out of the entire pool of aircraft combinations, only about 15 percent on average had a chance of being in conflict.

Of the average 18,500 aircraft pairs per simulation that passed the gross filter, an average of 72 had conflicts, representing about 0.06 percent of the original pool of 120,000 aircraft. Out of the nine simulation runs, only one missed alert was found, making the grand average missed alert probability equal to 0.0016 (see Tables 3.1-3, 3.1-4, 3.1-5 and 3.1-6). The strategic missed alert probability, which expands upon the missed alert definition to include valid alerts with a warning time of less than five minutes, is higher at 0.0408 (again referring to Tables 3.1-3, 3.1-4, 3.1-5 and 3.1-6). False alert probability is substantially higher at an aggregate measure (i.e., both red and yellow alerts) of 0.8481 (see Tables 3.1-3, 3.1-7, 3.1-8 and 3.1-9). In other words, for the combination of nine simulation runs, the conditional probability that a given URET alert does not have a matching actual standard separation violation is 0.8481. Finally, the grand average warning time is estimated at 15 minutes (see Table 3.1-10) and the average conflict start time delta (the absolute value time difference between the URET-predicted conflict start time and the actual conflict start time) is 222.38 seconds (3.7 minutes; see Table 3.1-11).

Grand Totals For All Simulation Runs:	
Scenario Information	
Total number of aircraft	4412
Total number of aircraft pair combinations	1083010
Total number of aircraft pairs with time overlap	281530
Alert Information	
Total URET Alert Count	4270
Total URET Red Alert Count	1644
Total URET Yellow Alert Count	2626
Total Valid Alert Count	647
Total Missed Alert Count	1
Total Strategic Missed Alert Count	27
Total False Alert Count	3623
Conflict Information	
Total count of track versus track conflicts	648
Total count of aircraft passing gross conflict filter	166402

Table 3.1-1: Summary Table Grand Totals for Analysis A

Grand Averages For All Simulation Runs:	
Scenario Information	
Average number of aircraft (per simulation run)	490.22
Average number of aircraft pair combinations	120334.44
Average number of aircraft pairs with time overlap	31281.11
Alert Information	
Average URET Alert Count	474.44
Average URET Red Alert Count	182.67
Average URET Yellow Alert Count	291.78
Average Valid Alert Count	71.89
Average Missed Alert Count	0.11
Average Strategic Missed Alert Count	3
Average False Alert Count	402.56
Conflict Information	
Average count of track versus track conflicts	72
Average count of aircraft passing gross conflict filter	18489.11

Table 3.1-2: Summary Table Grand Averages for Analysis A

Aggregate Missed Alert Probability			
Grand Average Missed Alert Probability			0.0016
Confidence Interval for Missed Alert Probability			
Upper Limit			0.0047
Lower Limit			0
Aggregate Strategic Missed Alert Probability			
Grand Average Strategic Missed Alert Probability			0.0408
Confidence Interval for Strategic Missed Alert Probability			
Upper Limit			0.053
Lower Limit			0.0286
Aggregate False Alert Probability			
	URET Alert Level:		
	Red	Yellow	Total
Grand Average FA Counts	132.11	270.44	402.56
Grand Average False Alert Probabilities	0.278	0.5701	0.8481
Confidence Interval for False Alert Probability			
Upper Limit	0.2914	0.5871	0.8550
Lower Limit	0.2646	0.5532	0.8413

Table 3.1-3: Aggregate Missed Alert, Strategic Missed Alert, and False Alert Probabilities[2]

[2] The grand average probabilities calculated in this table include some small roundoff error.

3.1.1 Missed Alert and Strategic Missed Alert Statistics

As shown in Table 3.1-3, the aggregate average missed alert probability was around 0.002 and the confidence interval limits ranged from 0.005 to 0. The important result here is that we cannot reject the claim that on average the probability of a missed alert is zero, since the confidence interval contains zero. The strategic missed alert (SMA) probability is another case altogether. For a strategic missed alert, the conflict prediction is required to have a warning time of 5 minutes minimum. The average SMA probability was around 0.04 and the confidence interval limits ranged from 0.05 to 0.03.

The averages, upper confidence limits, and lower confidence limits for the missed alert and strategic missed alert probabilities are presented in Table 3.1-4 to Table 3.1-6, respectively. These tables partition the missed alert probabilities by three factors: the minimum horizontal separation for the duration of the conflict, the vertical separation at this horizontal separation encounter, and whether the altitude at the minimum horizontal encounter of both aircraft was above or below flight level 290.

Missed Alert Conditional Probability			
	Altitude at h $\leq$ FL290	Altitude at h > FL290	
	$0 \leq v < 1$	$0 \leq v < 1$	$1 \leq v < 2$
$0 \leq h < 3$	0	0.0016	0
$3 \leq h < 5$	0	0	0
Strategic Missed Alert Conditional Probability			
	Altitude at h $\leq$ FL290	Altitude at h > FL290	
	$0 \leq v < 1$	$0 \leq v < 1$	$1 \leq v < 2$
$0 \leq h < 3$	0.013	0.0147	0.0014
$3 \leq h < 5$	0.0012	0.0074	0.0031

Note: h = minimum horizontal separation distance in nautical miles for the duration of the conflict;
v = vertical separation at minimum horizontal encounter in 1000's feet;
Altitude at h = maximum altitude of both aircraft in conflict at the time of minimum horizontal separation

Table 3.1-4: Analysis A Partitioned Average Missed Alert and Strategic Missed Alert Probability

Missed Alert Conditional Probability			
	Altitude at h $\leq$ FL290	Altitude at h > FL290	
	$0 \leq v < 1$	$0 \leq v < 1$	$1 \leq v < 2$
$0 \leq h < 3$	0	0.0046	0
$3 \leq h < 5$	0	0	0
Strategic Missed Alert Conditional Probability			
	Altitude at h $\leq$ FL290	Altitude at h > FL290	
	$0 \leq v < 1$	$0 \leq v < 1$	$1 \leq v < 2$
$0 \leq h < 3$	0.0187	0.0229	0.004
$3 \leq h < 5$	0.0035	0.0133	0.0069

Note: h = minimum horizontal separation distance in nautical miles for the duration of the conflict;
v = vertical separation at minimum horizontal encounter in 1000's feet;
Altitude at h = maximum altitude of both aircraft in conflict at the time of minimum horizontal separation

Table 3.1-5: Analysis A Upper 90% Confidence Limits for the Partitioned Average Missed Alert and Strategic Missed Alert Probability

Missed Alert Conditional Probability			
	Altitude at h $\leq$ FL290	Altitude at h > FL290	
	$0 \leq v < 1$	$0 \leq v < 1$	$1 \leq v < 2$
$0 \leq h < 3$	0	0	0
$3 \leq h < 5$	0	0	0
Strategic Missed Alert Conditional Probability			
	Altitude at h $\leq$ FL290	Altitude at h > FL290	
	$0 \leq v < 1$	$0 \leq v < 1$	$1 \leq v < 2$
$0 \leq h < 3$	0.0073	0.0065	0
$3 \leq h < 5$	0	0.0015	0

Note: h = minimum horizontal separation distance in nautical miles for the duration of the conflict;
v = vertical separation at minimum horizontal encounter in 1000's feet;
Altitude at h = maximum altitude of both aircraft in conflict at the time of minimum horizontal separation

Table 3.1-6: Analysis A Lower 90% Confidence Limits for the Partitioned Average Missed Alert and Strategic Missed Alert Probability

3.1.2 False Alert Statistics

As indicated in Table 3.1-3, the aggregate average false alert probability for Analysis A is around 0.85. This means that, given URET predicts an alert, 85 percent of the time it will be false or have no actual violation of standard separation. The confidence interval ranges between 0.855 and 0.841. Also the false alert probability is partitioned by alert level (e.g., the probability on average that a given alert is both false and red is approximately 0.28). The false alert probability can also be conditioned by specific alert level (i.e., red or yellow). This is determined by dividing the grand average FA count for a particular alert level by the average count of all alerts of that level. When the probabilities are conditioned in this manner for Analysis A, the conditional probability for a red alert being false is 0.72 (i.e., 132/183, as shown in Tables 3.1-2 and 3.1-3). In other words, given a red alert, the probability that the alert is false is 0.72. Similarly, given a yellow alert, the probability that the alert is false is 0.93 (i.e., 270/292, as shown in Tables 3.1-2 and 3.1-3).

Tables 3.1-7 through 3.1-9 present the false alert probabilities for both red and yellow alerts partitioned by three factors: the minimum horizontal separation between the aircraft's track position data, the vertical separation at this minimum horizontal separation encounter, and whether the altitude at the minimum horizontal encounter of both aircraft was above or below flight level 290. Table 3.1-7 presents the average false alert probabilities and Tables 3.1-8 and 3.1-9 present the upper and lower confidence interval limits. (Note: These tables do not contain the false alert probabilities for alerts which occur at horizontal separation greater than 30 nautical miles or vertical separation greater than 5000 feet. The counts for these events are contained in Table A.0-2 in Appendix A.)

For alerts with encounter altitudes below flight level 290, one would expect zero probabilities of false alerts for minimum horizontal distances between zero and five nautical miles, and vertical separation between zero and 1000 feet, since by definition these separations imply a standard separation violation. However, there exist small false alert probabilities in Table 3.1-7. As described in section 2.3.1, there are two exceptions applied for separation violations: either the duration of the conflict is smaller than a parameter threshold (e.g. 10 seconds), or both aircraft are in cruise flight and exhibiting a minimum vertical separation up to but not including 300 feet less than standard vertical separation. These two exceptions appear to be the cause of these rare cases of false alerts for encounter altitudes below flight level 290. There is an analogous situation for the encounter altitudes above flight level 290 shown in Table 3.1-7.

	Altitude at Horizontal Encounter $\leq$ FL290												
	$0 \leq v < 1$		$1 \leq v < 2$		$2 \leq v < 3$		$3 \leq v < 4$		$4 \leq v < 5$		Row Sub-Totals		Totals
	R	Y	R	Y	R	Y	R	Y	R	Y	R	Y	
$0 \leq h < 5$	0.002	0.002	0.008	0.005	0.007	0.008	0.004	0.006	0.005	0.004	0.026	0.025	0.051
$5 \leq h < 10$	0.005	0.019	0.002	0.014	0.002	0.01	0.002	0.01	0.002	0.005	0.013	0.058	0.071
$10 \leq h < 15$	0.002	0.006	0	0.004	0	0.006	0.001	0.003	0	0.002	0.003	0.021	0.024
$15 \leq h < 20$	0	0.003	0	0.002	0.001	0.001	0	0.001	0	0.001	0.001	0.008	0.009
$20 \leq h < 25$	0	0.001	0	0.001	0	0	0	0	0	0	0	0.002	0.002
$25 \leq h < 30$	0.001	0	0.001	0	0	0	0	0	0	0	0.002	0	0.002
Sub-Totals	0.01	0.031	0.011	0.026	0.01	0.025	0.007	0.02	0.007	0.012			
Totals	0.041		0.037		0.035		0.027		0.019				

	Altitude at Horizontal Encounter > FL290												
	$0 \leq v < 1$		$1 \leq v < 2$		$2 \leq v < 3$		$3 \leq v < 4$		$4 \leq v < 5$		Row Sub-Totals		Totals
	R	Y	R	Y	R	Y	R	Y	R	Y	R	Y	
$0 \leq h < 5$	0.009	0.006	0.011	0.005	0.025	0.011	0.007	0.006	0.011	0.009	0.063	0.037	0.1
$5 \leq h < 10$	0.053	0.102	0.003	0.021	0.005	0.025	0.002	0.01	0.003	0.008	0.066	0.166	0.232
$10 \leq h < 15$	0.027	0.072	0.001	0.009	0.003	0.01	0.001	0.004	0.002	0.005	0.034	0.1	0.134
$15 \leq h < 20$	0.013	0.023	0.001	0.002	0.001	0.003	0.001	0.002	0.001	0.001	0.017	0.031	0.048
$20 \leq h < 25$	0.005	0.014	0	0.002	0.002	0.001	0.001	0.001	0	0.001	0.008	0.019	0.027
$25 \leq h < 30$	0.003	0.004	0	0	0	0	0	0	0	0.001	0.003	0.005	0.008
Sub-Totals	0.11	0.221	0.016	0.039	0.036	0.05	0.012	0.023	0.017	0.025			
Totals	0.331		0.055		0.086		0.035		0.042				

Note: h = minimum horizontal separation distance in nautical miles; v = vertical separation at minimum horizontal encounter in 1000's feet

Note: R = Red URET Alert, Y= Yellow URET Alert

Table 3.1-7: Analysis A Average Partitioned False Alert Probabilities

	Altitude at Horizontal Encounter $\leq$ FL290												
	$0 \leq v < 1$		$1 \leq v < 2$		$2 \leq v < 3$		$3 \leq v < 4$		$4 \leq v < 5$		Row Sub-Totals		Totals
	R	Y	R	Y	R	Y	R	Y	R	Y	R	Y	
$0 \leq h < 5$	0.004	0.003	0.012	0.008	0.009	0.011	0.006	0.009	0.007	0.006	0.038	0.037	0.075
$5 \leq h < 10$	0.006	0.022	0.003	0.018	0.003	0.012	0.003	0.013	0.003	0.008	0.018	0.073	0.091
$10 \leq h < 15$	0.003	0.01	0	0.005	0.001	0.008	0.002	0.004	0.001	0.003	0.007	0.03	0.037
$15 \leq h < 20$	0.001	0.004	0.001	0.004	0.001	0.002	0.001	0.002	0	0.002	0.004	0.014	0.018
$20 \leq h < 25$	0.001	0.002	0	0.002	0	0.001	0	0	0	0.001	0.001	0.006	0.007
$25 \leq h < 30$	0.001	0.001	0.001	0	0	0.001	0	0	0	0.001	0.002	0.003	0.005
Sub-Totals	0.016	0.042	0.017	0.037	0.014	0.035	0.012	0.028	0.011	0.021			
Totals	0.058		0.054		0.049		0.04		0.032				

	Altitude at Horizontal Encounter > FL290												
	$0 \leq v < 1$		$1 \leq v < 2$		$2 \leq v < 3$		$3 \leq v < 4$		$4 \leq v < 5$		Row Sub-Totals		Totals
	R	Y	R	Y	R	Y	R	Y	R	Y	R	Y	
$0 \leq h < 5$	0.011	0.009	0.013	0.007	0.028	0.015	0.009	0.009	0.013	0.014	0.074	0.054	0.128
$5 \leq h < 10$	0.058	0.114	0.005	0.026	0.008	0.033	0.003	0.013	0.005	0.01	0.079	0.196	0.275
$10 \leq h < 15$	0.03	0.083	0.002	0.012	0.005	0.016	0.002	0.007	0.004	0.008	0.043	0.126	0.169
$15 \leq h < 20$	0.018	0.028	0.003	0.004	0.002	0.004	0.002	0.003	0.001	0.003	0.026	0.042	0.068
$20 \leq h < 25$	0.007	0.017	0	0.003	0.003	0.002	0.001	0.001	0.001	0.002	0.012	0.025	0.037
$25 \leq h < 30$	0.004	0.007	0	0.001	0	0.001	0	0.001	0	0.001	0.004	0.011	0.015
Sub-Totals	0.128	0.258	0.023	0.053	0.046	0.071	0.017	0.034	0.024	0.038			
Totals	0.386		0.076		0.117		0.051		0.062				

Note: h = minimum horizontal separation distance in nautical miles; v = vertical separation at minimum horizontal encounter in 1000's feet
Note: R = Red URET Alert, Y= Yellow URET Alert

Table 3.1-8: Analysis A Upper 90% Confidence Limits for the Average Partitioned False Alert Probabilities

	Altitude at Horizontal Encounter $\leq$ FL290												
	$0 \leq v < 1$		$1 \leq v < 2$		$2 \leq v < 3$		$3 \leq v < 4$		$4 \leq v < 5$		Row Sub-Totals		Totals
	R	Y	R	Y	R	Y	R	Y	R	Y	R	Y	
$0 \leq h < 5$	0.001	0.001	0.005	0.003	0.004	0.005	0.003	0.003	0.003	0.003	0.016	0.015	0.031
$5 \leq h < 10$	0.003	0.015	0.001	0.01	0	0.008	0.001	0.007	0	0.003	0.005	0.043	0.048
$10 \leq h < 15$	0.001	0.002	0	0.002	0	0.005	0	0.001	0	0.001	0.001	0.011	0.012
$15 \leq h < 20$	0	0.001	0	0.001	0	0	0	0	0	0	0	0.002	0.002
$20 \leq h < 25$	0	0.001	0	0	0	0	0	0	0	0	0	0.001	0.001
$25 \leq h < 30$	0	0	0	0	0	0	0	0	0	0	0	0	0
Sub-Totals	0.005	0.02	0.006	0.016	0.004	0.018	0.004	0.011	0.003	0.007			
Totals	0.025		0.022		0.022		0.015		0.01				

	Altitude at Horizontal Encounter > FL290												
	$0 \leq v < 1$		$1 \leq v < 2$		$2 \leq v < 3$		$3 \leq v < 4$		$4 \leq v < 5$		Row Sub-Totals		Totals
	R	Y	R	Y	R	Y	R	Y	R	Y	R	Y	
$0 \leq h < 5$	0.006	0.004	0.009	0.003	0.022	0.007	0.005	0.004	0.009	0.005	0.051	0.023	0.074
$5 \leq h < 10$	0.047	0.089	0.002	0.015	0.003	0.017	0	0.007	0.001	0.007	0.053	0.135	0.188
$10 \leq h < 15$	0.023	0.06	0.001	0.005	0.001	0.005	0	0.002	0	0.003	0.025	0.075	0.1
$15 \leq h < 20$	0.009	0.018	0	0.001	0	0.002	0	0.001	0	0	0.009	0.022	0.031
$20 \leq h < 25$	0.003	0.01	0	0.001	0	0.001	0	0	0	0	0.003	0.012	0.015
$25 \leq h < 30$	0.002	0.002	0	0	0	0	0	0	0	0	0.002	0.002	0.004
Sub-Totals	0.09	0.183	0.012	0.025	0.026	0.032	0.005	0.014	0.01	0.015			
Totals	0.273		0.037		0.058		0.019		0.025				

Note: h = minimum horizontal separation distance in nautical miles; v = vertical separation at minimum horizontal encounter in 1000's feet
Note: R = Red URET Alert, Y= Yellow URET Alert

Table 3.1-9: Analysis A Lower 90% Confidence Limits for the Average Partitioned False Alert Probabilities

3.1.3 Conflict Notification Timeliness Statistics

As shown in Table 3.1-10, the grand average warning time is around 15 minutes on average, with a confidence interval ranging from 15 minutes 25 seconds to 14 minutes 37 seconds. Since we match alerts with conflicts by taking the first alert presented to the URET display (i.e., notified to a controller), alerts can have relatively large positive warning times (e.g., average maximum warning time was around 38 minutes). For example, an alert may be predicted based on a trajectory built only from flight plan data using the estimated coordination fix time. The predicted conflict start time for this alert would then be based on the aircraft's position at the estimated coordination fix time. If the actual coordination fix time is later than the estimated time upon which the conflict prediction was based, this could cause the notification time to be more than the expected twenty minutes. In other words, URET makes a conflict prediction based on a trajectory that is modeling the aircraft longitudinally **ahead** of what actually happens, resulting in a warning time greater than 20 minutes.

Conflict start time delta, as discussed in Section 2.2.2, is the absolute value difference between the URET-predicted start time of the conflict and the actual conflict start time for valid alerts. On average the conflict start time delta was 222 seconds or 3.7 minutes. The confidence interval limits for the average conflict start time delta ranged from 264 to 180 seconds. This statistic is directly influenced by which alert is matched as the valid alert. Once again, the alert matched was the first notified to the controller, which may or may not have been predicted with a trajectory that had already synchronized with track data. The design of the matching rule was a trade off between warning time and conflict start time delta. If a later alert was chosen as the valid alert, perhaps the conflict start time delta would be less, but the warning time would be reduced as well, giving less warning for an impending conflict situation.

Conflict Notification Timeliness (seconds)	
Grand Average Warning Time	900.92
Grand Standard Deviation of Warning Time	361.89
Average Maximum Warning Time	2267.9
Average Minimum Warning Time	329.56
Average Valid Alerts Used	50.11
Confidence Interval for Warning Time	
Upper Limit	924.45
Lower Limit	877.39

Table 3.1-10: Analysis A Conflict Notification Timeliness Statistics

Conflict start time delta (delta of predicted vs. actual in seconds)	
Grand Average Conflict Start Time Delta	222.38
Grand Standard Deviation Conflict Start Time Delta	383.14
Average Maximum Conflict Start Time Delta	2053.2
Average Minimum Conflict Start Time Delta	5.33
Average Valid Alerts Used	50.11
Confidence Interval for Conflict Start Time Delta	
Upper Limit	264.55
Lower Limit	180.22

Table 3.1-11: Analysis A Conflict Start Time Delta Statistics

3.2 Analysis B

As stated in Section 3, Analysis B uses the same scenario and URET alerts as Analysis A but defines the conflicts with an expanded horizontal separation (i.e., 10 nautical miles). This shift from the standard definition of a conflict is worthwhile because a conflict probe does not strictly predict violations of standard separations, but predicts encounters at distances often much larger than standard separation. URET presents yellow alerts at a predicted horizontal separation distance of 10 nautical miles. By comparing URET against actual conflicts with an expanded 10 nautical mile separation threshold, the false alerts previously counted in Analysis A should appreciably be reduced. However, expanding the conflict separation thresholds increases the likelihood of a missed alert. This is exactly what is shown in the following tables of results.

Tables 3.2-1 and Table 3.2-2 present the scenario information and various alert events (i.e., missed alert, false alert, strategic missed alert) for the Analysis B processing of the nine simulation runs in terms of grand totals and grand average values. As expected, Analysis B has the same number of aircraft and the same number of aircraft with time overlap and passing the gross filter. The major difference is that the average number of conflicts increased from around 72 in Analysis A to 159 in Analysis B. That is, the percentage of aircraft in conflict from the pool of 120,000 aircraft combination pairs was around 0.13 compared to 0.06 in Analysis A. This is certainly interesting that doubling the horizontal separation to 10 nautical miles yields roughly twice the number of conflicts, on average. Out of the nine simulation runs processed for Analysis B, 13 missed alerts were found, making the grand average missed alert probability equal to 0.0092 (see Tables 3.2-3, 3.2-4, 3.2-5 and 3.2-6). The strategic missed alert probability, which expands upon the missed alert definition to include valid alerts with a warning time of less than five minutes, is slightly higher at 0.0696 (again referring to Tables 3.2-3, 3.2-4, 3.2-5 and 3.2-6). False alert probability is substantially lower at an aggregate measure (i.e., both red and yellow alerts) of 0.6451 (see Tables 3.2-3, 3.2-7, 3.2-8 and 3.2-9). Finally, the grand average warning time is estimated at 13.5 minutes (see Table 3.2-10) and the average conflict start time delta is 130.46 seconds (2 minutes and 10 seconds; see Table 3.2-11).

As shown in Tables 3.2-1 and 3.2-2, the average number of alerts decreased from around 475 to 444 for Analysis B. The actual URET alerts did not change, rather the difference is a result of the CAT exclusion rules applied to alerts. That is, as discussed in Section 2.3.2, duplicate valid alerts are excluded from the alert count, and since there are more valid alerts for Analysis B (i.e., 158 on average compared to 72) it is expected that we would exclude more duplicate alerts from the analysis. Also, as expected, the average number of false alerts decreased from 403 to 286, which is a reduction of around 30 percent. Since Analysis B considers violations of horizontal separations between 5 and 10 nautical miles to be conflicts, these encounters are now considered valid alerts resulting in a reduction of the false alert count.

Grand Totals For All Simulation Runs:	
Scenario Information	
Total number of aircraft	4412
Total number of aircraft pair combinations	1083010
Total number of aircraft pairs with time overlap	281530
Alert Information	
Total URET Alert Count	3993
Total URET Red Alert Count	1564
Total URET Yellow Alert Count	2429
Total Valid Alert Count	1420
Total Missed Alert Count	13
Total Strategic Missed Alert Count	100
Total False Alert Count	2573
Conflict Information	
Total count of track versus track conflicts	1433
Total count of aircraft passing gross conflict filter	166402

Table 3.2-1: Summary Table Grand Totals for Analysis B

Grand Averages For All Simulation Runs:	
Scenario Information	
Average number of aircraft (per simulation run)	490.22
Average number of aircraft pair combinations	120334.44
Average number of aircraft pairs with time overlap	31281.11
Alert Information	
Average URET Alert Count	443.67
Average URET Red Alert Count	173.78
Average URET Yellow Alert Count	269.89
Average Valid Alert Count	157.78
Average Missed Alert Count	1.44
Average Strategic Missed Alert Count	11.11
Average False Alert Count	285.89
Conflict Information	
Average count of track versus track conflicts	159.22
Average count of aircraft passing gross conflict filter	18489.11

Table 3.2-2: Summary Table Grand Averages for Analysis B

Aggregate Missed Alert Probability			
Grand Average Missed Alert Probability			0.0092
Confidence Interval for Missed Alert Probability			
Upper Limit			0.0133
Lower Limit			0.0051
Aggregate Strategic Missed Alert Probability			
Grand Average Strategic Missed Alert Probability			0.0696
Confidence Interval for Strategic Missed Alert Probability			
Upper Limit			0.0781
Lower Limit			0.0612
Aggregate False Alert Probability			
	URET Alert Level:		
	Red	Yellow	Total
Grand Average FA Counts	88.44	197.44	285.89
Grand Average False Alert Probabilities	0.1992	0.4458	0.6451
Confidence Interval for False Alert Probability			
Upper Limit	0.2103	0.4606	0.6569
Lower Limit	0.1881	0.4311	0.6332

Table 3.2-3: Aggregate Missed, Strategic Missed, and False Alert Probabilities[3]

3.2.1 Missed Alert and Strategic Missed Alert Statistics

The missed alert statistics in this analysis are based on URET not detecting a conflict with the expanded (i.e., ten nautical mile) separation standard. As shown in Table 3.2-3, the average missed alert probability for Analysis B was around 0.009, with confidence intervals ranging from 0.013 to 0.005. In other words, with 90 percent confidence the average missed alert probability increased by 0.007. This is a substantial increase (more than double) from Analysis A. The important point here is that the Analysis A average missed alert probability of 0.0016 is not even within the Analysis B confidence interval. A possible explanation for this is discussed in section 3.3.

The increase in separation distance for Analysis B not only effects the strategic missed alert probability, but also creates conflicts that occur earlier. If the conflicts start earlier, the warning time will be reduced, consequently increasing the probability of strategic missed alerts (since SMAs include valid alerts that have a warning time of less than 5 minutes).

The averages, upper confidence limits, and lower confidence limits for the missed alert and strategic missed alert probabilities are presented in Tables 3.2-4, 3.2-5 and 3.2-6. As in Analysis A, these tables are partitioned by three factors: minimum horizontal separation for the duration of the conflict, the vertical separation at this horizontal separation encounter, and whether the altitude at the minimum horizontal encounter of both aircraft was above or below flight level 290.

[3] The grand average probabilities calculated in this table include some small roundoff error.

Missed Alert Conditional Probability			
	Altitude at h $\leq$ FL290	Altitude at h > FL290	
	$0 \leq v < 1$	$0 \leq v < 1$	$1 \leq v < 2$
$0 \leq h < 3$	0.0008	0.0007	0
$3 \leq h < 5$	0	0	0.0008
$5 \leq h < 10$	0.0029	0.0041	0
Strategic Missed Alert Conditional Probability			
	Altitude at h $\leq$ FL290	Altitude at h > FL290	
	$0 \leq v < 1$	$0 \leq v < 1$	$1 \leq v < 2$
$0 \leq h < 3$	0.0067	0.008	0
$3 \leq h < 5$	0.0011	0.0099	0.0027
$5 \leq h < 10$	0.0086	0.0271	0.0055

Note: h = minimum horizontal separation distance in nautical miles for the duration of the conflict;
v = vertical separation at minimum horizontal encounter in 1000's feet;
Altitude at h = maximum altitude of both aircraft in conflict at the time of minimum horizontal separation

Table 3.2-4: Analysis B Partitioned Average Missed Alert and Strategic Missed Alert Probability

Missed Alert Conditional Probability			
	Altitude at h $\leq$ FL290	Altitude at h > FL290	
	$0 \leq v < 1$	$0 \leq v < 1$	$1 \leq v < 2$
$0 \leq h < 3$	0.0022	0.002	0
$3 \leq h < 5$	0	0	0.0022
$5 \leq h < 10$	0.005	0.0066	0
Strategic Missed Alert Conditional Probability			
	Altitude at h $\leq$ FL290	Altitude at h > FL290	
	$0 \leq v < 1$	$0 \leq v < 1$	$1 \leq v < 2$
$0 \leq h < 3$	0.0105	0.0119	0
$3 \leq h < 5$	0.0032	0.0158	0.0052
$5 \leq h < 10$	0.0109	0.0345	0.0103

Note: h = minimum horizontal separation distance in nautical miles for the duration of the conflict;
v = vertical separation at minimum horizontal encounter in 1000's feet;
Altitude at h = maximum altitude of both aircraft in conflict at the time of minimum horizontal separation

Table 3.2-5: Analysis B Upper 90% Confidence Limits for the Partitioned Average Missed Alert and Strategic Missed Alert Probability

Missed Alert Conditional Probability			
	Altitude at h $\leq$ FL290	Altitude at h > FL290	
	$0 \leq v < 1$	$0 \leq v < 1$	$1 \leq v < 2$
$0 \leq h < 3$	0	0	0
$3 \leq h < 5$	0	0	0
$5 \leq h < 10$	0.0008	0.0016	0
Strategic Missed Alert Conditional Probability			
	Altitude at h $\leq$ FL290	Altitude at h > FL290	
	$0 \leq v < 1$	$0 \leq v < 1$	$1 \leq v < 2$
$0 \leq h < 3$	0.0029	0.0041	0
$3 \leq h < 5$	0	0.004	0.0002
$5 \leq h < 10$	0.0063	0.0197	0.0007

Note: h = minimum horizontal separation distance in nautical miles for the duration of the conflict;
v = vertical separation at minimum horizontal encounter in 1000's feet;
Altitude at h = maximum altitude of both aircraft in conflict at the time of minimum horizontal separation

Table 3.2-6: Analysis B Lower 90% Confidence Limits for the Partitioned Average Missed Alert and Strategic Missed Alert Probability

3.2.2 False Alert Statistics

As indicated in Table 3.2-3, the aggregate average false alert probability for Analysis B is 0.6451, which is 0.20 less than Analysis A. That is, for Analysis B given there is a URET alert, 65 percent of the time it will be false or have no actual conflict of expanded separation. The confidence interval ranges from 0.657 to 0.633. Also the false alert probability is partitioned by alert level. For example, the probability on average that a given alert is both false and red is around 0.199. As discussed in Section 3.2.1, the false alert probability can also be conditioned by specific alert level. When the probabilities are conditioned in this manner for Analysis B, the conditional probability for a red alert being false is 0.51 (i.e., 88/174, as shown in Tables 3.2-2 and 3.2-3). In other words, given a red alert, the probability that the alert is false is 0.51. Similarly, given a yellow alert, the probability that the alert is false is 0.73 (i.e., 198/270, as shown in Tables 3.2-2 and 3.2-3).

As in Analysis A, Tables 3.2-7 through 3.2-9 present the Analysis B false alert probabilities for both red and yellow alerts partitioned by three factors: the minimum horizontal separation between the aircraft's track position data, the vertical separation at this minimum horizontal separation encounter, and whether the altitude at the minimum horizontal encounter of both aircraft was above or below flight level 290. First, Table 3.2-7 presents the average false alert probabilities, and then Tables 3.2-8 and 3.2-9 present the upper and lower confidence interval limits. (Note: These tables do not contain the false alert probabilities for alerts which occur at horizontal separation greater than 30 nautical miles or vertical separation greater than 5000 feet. The counts for these events are contained in Table B.0-2 in Appendix B.)

	Altitude at Horizontal Encounter ≤ FL290												
	$0 \leq v < 1$		$1 \leq v < 2$		$2 \leq v < 3$		$3 \leq v < 4$		$4 \leq v < 5$		Row Sub-Totals		Totals
	R	Y	R	Y	R	Y	R	Y	R	Y	R	Y	
$0 \leq h < 5$	0.001	0.002	0.004	0.004	0.005	0.007	0.004	0.006	0.004	0.004	0.018	0.023	0.041
$5 \leq h < 10$	0.001	0.005	0.001	0.011	0.001	0.01	0.001	0.008	0.002	0.006	0.006	0.04	0.046
$10 \leq h < 15$	0.002	0.006	0	0.004	0	0.007	0.001	0.003	0	0.002	0.003	0.022	0.025
$15 \leq h < 20$	0.001	0.003	0	0.003	0.001	0.001	0	0.001	0	0.001	0.002	0.009	0.011
$20 \leq h < 25$	0	0.002	0	0.001	0	0	0	0	0	0	0	0.003	0.003
$25 \leq h < 30$	0.001	0	0.001	0	0	0	0	0	0	0	0.002	0	0.002
Sub-Totals	0.006	0.018	0.006	0.023	0.007	0.025	0.006	0.018	0.006	0.013			
Totals	0.024		0.029		0.032		0.024		0.019				

	Altitude at Horizontal Encounter > FL290												
	$0 \leq v < 1$		$1 \leq v < 2$		$2 \leq v < 3$		$3 \leq v < 4$		$4 \leq v < 5$		Row Sub-Totals		Totals
	R	Y	R	Y	R	Y	R	Y	R	Y	R	Y	
$0 \leq h < 5$	0	0	0.011	0.004	0.019	0.008	0.006	0.005	0.011	0.009	0.047	0.026	0.073
$5 \leq h < 10$	0.001	0.009	0.004	0.014	0.004	0.021	0.002	0.01	0.003	0.008	0.014	0.062	0.076
$10 \leq h < 15$	0.029	0.077	0.002	0.009	0.004	0.011	0.001	0.005	0.002	0.006	0.038	0.108	0.146
$15 \leq h < 20$	0.014	0.025	0.001	0.003	0.001	0.003	0.001	0.002	0.001	0.002	0.018	0.035	0.053
$20 \leq h < 25$	0.005	0.015	0	0.002	0.002	0.002	0.001	0.001	0	0.001	0.008	0.021	0.029
$25 \leq h < 30$	0.003	0.004	0	0	0	0	0	0	0	0.001	0.003	0.005	0.008
Sub-Totals	0.052	0.13	0.018	0.032	0.03	0.045	0.011	0.023	0.017	0.027			
Totals	0.182		0.05		0.075		0.034		0.044				

Note: h = minimum horizontal separation distance in nautical miles; v = vertical separation at minimum horizontal encounter in 1000's feet

Note: R = Red URET Alert, Y= Yellow URET Alert

Table 3.2-7: Analysis B Average Partitioned False Alert Probabilities

	Altitude at Horizontal Encounter ≤ FL290												
	0 ≤ v < 1		1 ≤ v < 2		2 ≤ v < 3		3 ≤ v < 4		4 ≤ v < 5		Row Sub-Totals		Totals
	R	Y	R	Y	R	Y	R	Y	R	Y	R	Y	
0 ≤ h < 5	0.001	0.003	0.006	0.007	0.007	0.01	0.006	0.008	0.006	0.006	0.026	0.034	0.06
5 ≤ h < 10	0.002	0.006	0.001	0.014	0.003	0.012	0.002	0.012	0.003	0.008	0.011	0.052	0.063
10 ≤ h < 15	0.003	0.01	0	0.005	0.001	0.009	0.002	0.005	0.001	0.003	0.007	0.032	0.039
15 ≤ h < 20	0.001	0.004	0.001	0.004	0.001	0.002	0.001	0.002	0	0.002	0.004	0.014	0.018
20 ≤ h < 25	0.001	0.003	0	0.002	0	0.001	0	0	0	0.001	0.001	0.007	0.008
25 ≤ h < 30	0.001	0.001	0.001	0	0	0.001	0	0	0	0.001	0.002	0.003	0.005
Sub-Totals	0.009	0.027	0.009	0.032	0.012	0.035	0.011	0.027	0.01	0.021			
Totals	0.036		0.041		0.047		0.038		0.031				

	Altitude at Horizontal Encounter > FL290												
	0 ≤ v < 1		1 ≤ v < 2		2 ≤ v < 3		3 ≤ v < 4		4 ≤ v < 5		Row Sub-Totals		Totals
	R	Y	R	Y	R	Y	R	Y	R	Y	R	Y	
0 ≤ h < 5	0	0	0.013	0.006	0.023	0.012	0.007	0.008	0.013	0.012	0.056	0.038	0.094
5 ≤ h < 10	0.002	0.013	0.005	0.017	0.007	0.027	0.003	0.012	0.005	0.01	0.022	0.079	0.101
10 ≤ h < 15	0.033	0.09	0.002	0.013	0.006	0.017	0.002	0.007	0.004	0.008	0.047	0.135	0.182
15 ≤ h < 20	0.019	0.03	0.003	0.004	0.002	0.005	0.002	0.003	0.002	0.003	0.028	0.045	0.073
20 ≤ h < 25	0.007	0.018	0	0.003	0.003	0.003	0.002	0.002	0.001	0.002	0.013	0.028	0.041
25 ≤ h < 30	0.005	0.007	0	0.001	0	0.001	0	0.001	0	0.001	0.005	0.011	0.016
Sub-Totals	0.066	0.158	0.023	0.044	0.041	0.065	0.016	0.033	0.025	0.036			
Totals	0.224		0.067		0.106		0.049		0.061				

Note: h = minimum horizontal separation distance in nautical miles; v = vertical separation at minimum horizontal encounter in 1000's feet
Note: R = Red URET Alert, Y= Yellow URET Alert

Table 3.2-8: Analysis B Upper 90% Confidence Limits for the Average Partitioned False Alert Probabilities

	Altitude at Horizontal Encounter $\leq$ FL290												
	$0 \leq v < 1$		$1 \leq v < 2$		$2 \leq v < 3$		$3 \leq v < 4$		$4 \leq v < 5$		Row Sub-Totals		Totals
	R	Y	R	Y	R	Y	R	Y	R	Y	R	Y	
$0 \leq h < 5$	0	0	0.001	0.001	0.003	0.004	0.003	0.003	0.002	0.003	0.009	0.011	0.02
$5 \leq h < 10$	0	0.003	0	0.008	0	0.007	0	0.005	0	0.003	0	0.026	0.026
$10 \leq h < 15$	0.001	0.002	0	0.002	0	0.005	0	0.001	0	0.001	0.001	0.011	0.012
$15 \leq h < 20$	0	0.002	0	0.001	0	0	0	0	0	0	0	0.003	0.003
$20 \leq h < 25$	0	0.001	0	0	0	0	0	0	0	0	0	0.001	0.001
$25 \leq h < 30$	0	0	0	0	0	0	0	0	0	0	0	0	0
Sub-Totals	0.001	0.008	0.001	0.012	0.003	0.016	0.003	0.009	0.002	0.007			
Totals	0.009		0.013		0.019		0.012		0.009				

	Altitude at Horizontal Encounter > FL290												
	$0 \leq v < 1$		$1 \leq v < 2$		$2 \leq v < 3$		$3 \leq v < 4$		$4 \leq v < 5$		Row Sub-Totals		Totals
	R	Y	R	Y	R	Y	R	Y	R	Y	R	Y	
$0 \leq h < 5$	0	0	0.009	0.002	0.015	0.005	0.004	0.002	0.008	0.005	0.036	0.014	0.05
$5 \leq h < 10$	0	0.006	0.002	0.01	0.001	0.015	0	0.007	0.001	0.007	0.004	0.045	0.049
$10 \leq h < 15$	0.025	0.064	0.001	0.006	0.002	0.005	0	0.002	0	0.003	0.028	0.08	0.108
$15 \leq h < 20$	0.01	0.019	0	0.001	0	0.002	0	0.001	0	0	0.01	0.023	0.033
$20 \leq h < 25$	0.003	0.011	0	0.001	0	0.001	0	0	0	0	0.003	0.013	0.016
$25 \leq h < 30$	0.002	0.002	0	0	0	0	0	0	0	0	0.002	0.002	0.004
Sub-Totals	0.04	0.102	0.012	0.02	0.018	0.028	0.004	0.012	0.009	0.015			
Totals	0.142		0.032		0.046		0.016		0.024				

Note: h = minimum horizontal separation distance in nautical miles; v = vertical separation at minimum horizontal encounter in 1000's feet
Note: R = Red URET Alert, Y= Yellow URET Alert

Table 3.2-9: Analysis B Lower 90% Confidence Limits for the Average Partitioned False Alert Probabilities

3.2.3 Conflict Notification Timeliness Statistics

As shown in Table 3.2-10, the grand average warning time for Analysis B is approximately 13.5 minutes on average, with a confidence interval ranging from 13 minutes 48 seconds to 13 minutes 7 seconds. As in Analysis A, URET alerts were matched with TCP-determined conflicts by taking the first alert presented to the URET display (i.e., notified to a controller), so alerts can have relatively large positive warning times (e.g. average maximum warning time was around 35.5 minutes). There is a reduction of about 1.5 minutes in the average warning time from Analysis A to B, since for Analysis B an aircraft pair will be in violation of a 10 nautical mile separation earlier than five nautical mile separation.

Conflict start time delta statistics (the absolute value difference between the predicted start time of the conflict and the actual start time for valid alerts as discussed in Section 2.2.2) are presented in Table 3.2-11. On average, the conflict start time delta was 130 seconds or 2 minutes 10 seconds. The confidence interval limits for the average conflict start time delta ranged from 158 to 102 seconds. It is interesting to note that, as expected, the average conflict start time delta decreased (92 seconds) from Analysis A to B. URET reports the conflict prediction start time based on violation of the conformance boxes plus five nautical miles separation; nominally this amounts to a 10 nautical mile separation. Therefore, it is expected that the conflict prediction start delta would be reduced for Analysis B, since the expansion of the separation standard for actual conflicts to 10 nautical miles is closer to what URET uses in predicting an aircraft-to-aircraft conflict.

Conflict Notification Timeliness (seconds)	
Grand Average Warning Time	808.07
Grand Standard Deviation of Warning Time	297.69
Average Maximum Warning Time	2124.44
Average Minimum Warning Time	310.22
Average Valid Alerts Used	95.11
Confidence Interval for Warning Time	
Upper Limit	828.73
Lower Limit	787.41

Table 3.2-10: Analysis B Conflict Notification Timeliness Statistics

Conflict Start Time Delta (delta of predicted vs. actual in seconds)	
Grand Average Conflict Start Time Delta	130.46
Grand Standard Deviation Conflict Start Time Delta	258.93
Average Maximum Conflict Start Time Delta	1789.56
Average Minimum Conflict Start Time Delta	0.78
Average Valid Alerts Used	95.11
Confidence Interval for Conflict Start Time Delta	
Upper Limit	158.74
Lower Limit	102.17

Table 3.2-11: Analysis B Conflict Start Time Delta Statistics

3.3 Observations

This section presents some detailed observations of the error events detected in Analyses A and B discussed in the previous two sections. Two MITRE/CAASD developed tools, AEC and Xeval, allowed ACT-250 to analyze the error events in detail and make these observations. The Xeval tool was used extensively to examine the URET predictions and compare these predictions to what was found by processing the HCS track data. These observations offer some explanation to why some of the error events took place, but they do not represent explanations for all the error events. In some cases, it is unknown why URET did not present an alert or present it in a timely manner. In general, the more serious error events (e.g., missed alerts) were very few making the manual task of examining them much easier.

3.3.1 Late Valid Alerts

In this study, strategic missed alerts are composed of both conflicts that do not have corresponding alerts and late valid alerts (LVAs). An LVA is defined to occur when URET predicts a conflict between two aircraft, and the actual conflict warning time is determined to be less than a parameter time (5 minutes) before the start of the conflict. The basic cause of an LVA seems to be an inaccurate prediction of one (or sometimes both) of the flight paths, or trajectory, at the time when URET should be predicting a conflict. Subsequently, URET corrects the flight path prediction by a reconformance and then predicts a conflict, but the prediction is too late to give an adequate warning time. This type of error was often observed in the simulation when one of the aircraft was either climbing or descending. URET was unable in these cases to predict accurately the flight path of a climbing aircraft or a descending aircraft sufficiently in advance of the start of the conflict.

The following sections discuss the circumstances under which LVAs can occur, and then examples of three LVAs which occurred in the simulation are described.

3.3.1.1 Conflict Geometries

The aircraft can be separated either horizontally or vertically or both. A conflict requires the simultaneous violation of the horizontal and vertical minimums. This can occur when (1) the aircraft flight paths cross and are at the same altitude, (2) the aircraft flight paths cross and one aircraft is descending or climbing through the other aircraft's altitude, (3) aircraft are flying the same route and altitude in trail and get too close, (4) aircraft are flying parallel routes at the same altitude, and (5) aircraft are flying parallel routes at the same altitude and one aircraft is descending or climbing through the other aircraft's altitude. Examples 1 and 2 in Section 3.3.1.5 show geometry (2); example 3 depicts geometry (5).

3.1.1.2 Altitude Transitions

The critical variable in modeling a climbing aircraft is its rate of climb; the critical variables in modeling a descending aircraft are its rate of descent and its top of descent (TOD). Errors can occur in predicting these variables. Examples 1 and 2 in Section 3.3.1.5 depict errors in predicting the TOD. These errors are often confounded with the fact that URET increases the conformance box dimensions in a descent or climb, which essentially reduce the reaction time of URET if aircraft track is outside these boxes. Track data would eventually be outside the conformance boxes if one of these occur, but an increased conformance box will delay URET from determining that a trajectory reconformance is necessary. This situation is depicted in Example 1 (see Section 3.3.1.5).

3.3.1.3 Trajectory Conformance to Track Reports

For climbing aircraft, URET often takes one to two minutes to reconform an aircraft trajectory after it starts receiving track reports on the aircraft. In a number of cases in the simulation, there was a conflict slightly over five minutes after the start of track data for a particular aircraft. In these instances, URET was unable to predict the conflict with a warning time of at least five minutes or more and an LVA resulted.

3.3.1.4 Alert Deletion

Sometimes URET will present an alert to a controller, maintain it for a few cycles, and then delete the alert before the predicted conflict occurs. This is an error when the conflict actually occurs. This type of error is not currently captured by our analysis and is related to URET's conflict prediction stability which will be investigated as part of future ACT-250 studies. In this study, an alert that is notified to a controller is counted as a valid alert regardless of whether it is deleted before the actual conflict start time.

3.3.1.5 Examples of Late Valid Alerts

Following is a discussion of three examples from the simulation runs which illustrate situations in which LVAs occurred.

Example 1

Aircraft A descends at the same time that its route crosses the route of Aircraft B. Initially both Aircraft A and Aircraft B are in level cruise, with Aircraft A 2000 feet above Aircraft B. Aircraft A's actual top of descent (TOD) is approximately two minutes later than its predicted TOD. As soon as URET thinks that Aircraft A is within a parameter distance of the predicted TOD, it expands its conformance box downward (The conformance box in level cruise extends from 300 feet below the nominal altitude of the aircraft to 300 feet above. When the aircraft starts a descent, the box is expanded to 1300 feet below the aircraft's presumed altitude.) As a result of this expansion, there is a delay in reconformance of the aircraft; that is, the aircraft is not reconformed for one minute after it starts its descent. During this time, Aircraft A reaches the cruise altitude of Aircraft B and it appears to URET that Aircraft A will safely descend below Aircraft B's altitude before Aircraft B comes close enough to be in conflict. Concurrently, the next radar track position supplied by the HCS for Aircraft A places it outside its conformance box and the aircraft is reconformed vertically. This reconformance places Aircraft A back at its cruise altitude and a conflict with Aircraft B is immediately predicted to occur in 127 seconds. In this example the error in predicting the TOD has reduced the warning time to two minutes, producing a late valid alert. Figure 3.3-1 depicts this situation for Aircraft A.

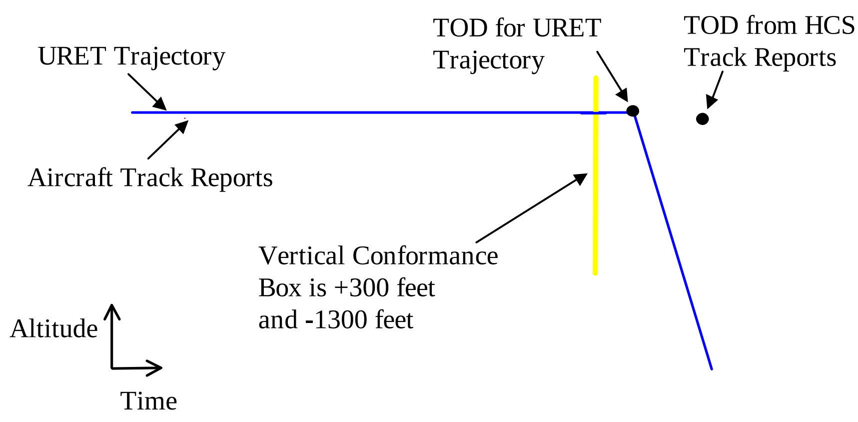


Figure 3.3-1: Time vs. Altitude Plot of Aircraft A's Trajectory and Track

Example 2

Two aircraft depart from two different ZID area airports. Aircraft A departs from Bloomington, Indiana and Aircraft B departs from Cincinnati, Ohio. Initially both aircraft are climbing to cruise altitude with crossing flight paths. Aircraft A levels off at 33,000 feet and Aircraft B continues to climb through this altitude. At the horizontal route crossing point, Aircraft A is in level cruise; Aircraft B is below Aircraft A and climbing. A conflict occurs before Aircraft A has crossed Aircraft B's path and continues after Aircraft B has crossed Aircraft A's path and is still below Aircraft A's cruise altitude. This situation in the horizontal dimension is depicted in Figure 3.3-2a.

The alert is delayed, providing a warning time of only 140 seconds for two reasons. First, the predicted longitudinal position of the cruising Aircraft A is in error; the URET alert appears when a longitudinal reconformance is done on Aircraft A placing it further along its trajectory and closer to Aircraft B. Secondly, the predicted climb gradient for Aircraft B is less than the simulated climb gradient. Aircraft B reaches Aircraft A's cruise altitude approximately 1.5 minutes sooner than predicted. The vertical reconformance due to the incorrect gradient is not done until after the conflict has ended due to the increasing horizontal separation of the aircraft. This is illustrated in Figure 3.3-2b. The effect of vertical reconformances on the accuracy of the conflict prediction is an area worthy of future study.

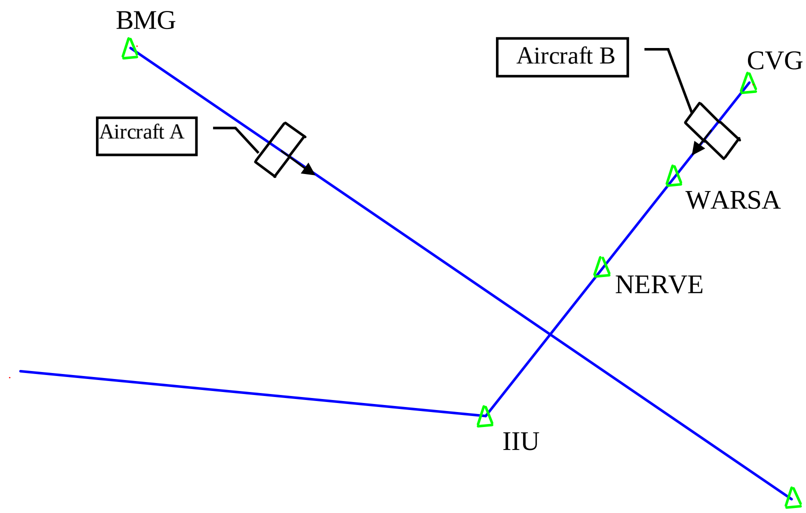


Figure 3.3-2a: Horizontal Plot of Aircraft A and Aircraft B (5 minutes before conflict start time)

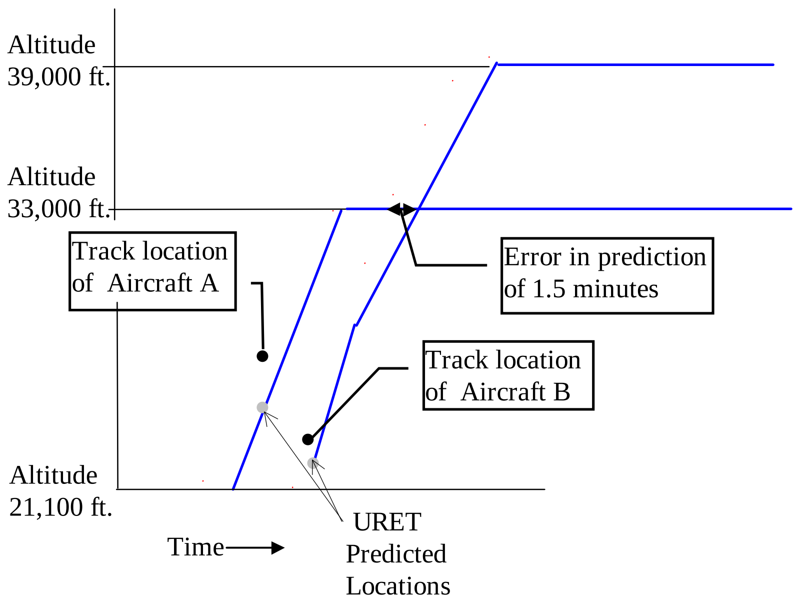


Figure 3.3-2b: Altitude vs. Time Plot of Aircraft A and Aircraft B (5 minutes before conflict start time)

Example 3

Aircraft A and Aircraft B are flying along on approximately parallel routes in opposite directions with Aircraft A 9000 feet higher than Aircraft B. Aircraft A descends, coming into conflict with Aircraft B when it passes through Aircraft B's altitude. Aircraft A's predicted TOD and predicted descent is later than its actual TOD and actual descent, and the predicted time of crossing Aircraft B's altitude is 82 seconds later than the actual time of altitude crossing. No conflict is predicted. As Aircraft A descends, a track report is received that places the aircraft outside of the top of the conformance box after about 50 seconds and 2000 feet of descent, causing a vertical reconformance. This vertical reconformance results in a new predicted flight path which places Aircraft A at Aircraft B's altitude 74 seconds before the actual altitude crossing time. No conflict is predicted. However, the new predicted flight path descends the aircraft at a steeper gradient than is simulated. Another vertical reconformance occurs after the aircraft descends an additional 3000 feet in 60 seconds. This second vertical reconformance predicts an altitude crossing 6 seconds later than the actual and results in a valid conflict prediction. The conflict prediction has occurred late and a warning time of only 97 seconds is provided. This example is illustrated in Figure 3.3-3.

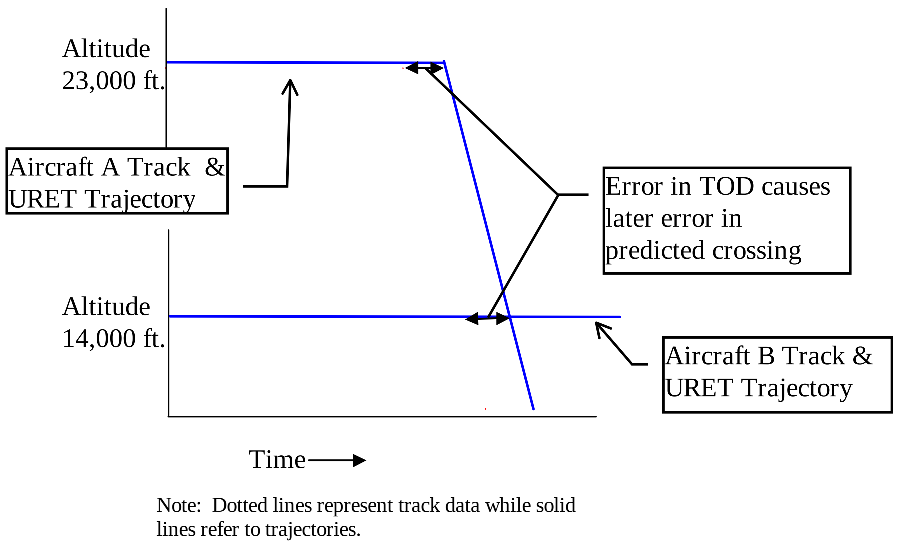


Figure 3.3-3: Altitude vs. Time Plot of Aircraft A and Aircraft B (5 minutes before conflict start time)

3.3.2 Missed Alerts

For Analysis A, one missed conflict detection was determined to have occurred during one of the nine simulation runs. The circumstances of the missed detection are described in the following paragraphs.

Aircraft A was in level cruise at 33,000 feet. Aircraft B was climbing on a crossing route from its departure airport to a cruise altitude of 35,000 feet. Aircraft B crosses Aircraft A's route while climbing, and during the climb comes into conflict with Aircraft A. At the start of the conflict, Aircraft B is at 33,400 feet climbing to 35,000 feet. Aircraft A is cruising at 32,900 feet, 2.87 nautical miles away from Aircraft B horizontally. The URET trajectory model has Aircraft B in level cruise at 35,000 feet. URET has track data (URET category A) for about 4 minutes before the conflict occurs. The URET trajectory never reconforms to match Aircraft B climbing track. After receiving track data URET reconforms the Aircraft B trajectory longitudinally, but URET does not make a vertical reconformance. Xeval shows Aircraft B to have a vertical drift count of 21 (when the aircraft is supposed to be in level flight according to its' flight plan and "drifts" vertically outside of the vertical conformance bounds, this is known as vertical drift; the count shown by Xeval represents the number of times the aircraft was determined to be in vertical drift). The horizontal trajectories match the horizontal tracks, so the prediction error is in the vertical and longitudinal dimensions only. The conflict lasts for 30 seconds. The reasons for this conflict not being detected are unknown. Misapplication of the vertical drift rule may be the cause.

URET predicts the minimum distance between the centers of the conformance boxes of two aircraft. When the aircraft stay at the centers of their conformance boxes, this prediction is the same as predicting the minimum separation of the aircraft (i.e., the trajectory centerline). Because the aircraft are usually not at the center of the conformance boxes, the actual minimum aircraft separation distance is either greater or less than the separation of the centers of the boxes. It sometimes happens that the first aircraft is near the boundary of its conformance box closest to the second aircraft, and at the same time the second aircraft is near the boundary of its conformance box which is closest to the first aircraft. When this happens, URET can correctly predict the separation of the conformance boxes, but it overestimates the minimum aircraft separation. This overestimate can and does cause missed alerts in the 5 to 10 nautical mile range. The URET design feature of making conformance boxes 5 nautical miles wide and requiring that the boxes be separated by 5 nautical miles guarantees, in so far as the predictions of the positions of the boxes are accurate, that the aircraft will be separated by a minimum horizontal distance of 5 nautical miles, even when the aircraft are at the edges of their conformance boxes.

Analysis B has several missed alerts in which the minimum horizontal separation of the aircraft is just less than 10 nautical miles. The conformance boxes were predicted to have adequate separation and in fact did appear to have adequate separation. However, the actual aircraft tracks within the boxes did not maintain a separation of 10 nautical miles.

3.3.3 False Alerts

The main reason for false alerts that was observed for both Analyses A and B was the distances URET uses to predict aircraft-to-aircraft conflicts. URET builds conformance boxes around the trajectory centerline and predicts if these conformance boxes violate separations (i.e., five nautical miles in the horizontal and 400 feet in the vertical below FL290, and 1400 feet above). The conformance boxes are nominally plus or minus 2.5 by 1.5 nautical miles in the horizontal dimension, and plus or minus 300 feet in the vertical dimension. Furthermore, the conformance boxes expand if the particular aircraft is in a transition either vertically or horizontally. The net effect is URET normally predicts conflicts of 10 nautical miles or greater. When comparing URET alerts against conflicts of standard separations, it would be expected that many of the URET alerts would not have corresponding standard separation conflicts. The simulation study did estimate the difference in false alert counts was around 25 percent considering conflicts at 5 and 10 nautical miles.

The fact that aircraft-to-aircraft encounters often occur during transitions adds to the URET conformance box sizes and increases the conflict separations it detects. This would increase the false alert probability even further if the encounters were still not within the actual horizontal distances of 5 or 10 nautical miles for Analysis A or B, respectively.

Further, URET initially builds trajectories based on only a flight plan (i.e., URET inbound category F). To build this initial trajectory URET uses the estimated coordination fix time, which may be in error. Therefore, the conflict predictions based on this initial trajectory may have relatively large longitudinal errors compared to the track versus track conflicts. In other words, this initial trajectory is yet another source of false alerts.

4. Summary

This report presents accuracy measures for URET's conflict prediction based on post-processing analysis of data collected during nine Indianapolis simulation runs of URET D3 in a single center operation. Seven ZID sectors were simulated (four high altitude and three low altitude) and a sample of three hours of SAR data of flights into and out of ZID was used as the basis for generating approximately 4500 simulated flights. Military flights and holding aircraft were excluded from the simulation, and all aircraft flew in adherence to their original filed flight plans without redirection due to controller actions. All procedural altitude restrictions were turned off in URET, and no winds were simulated. The results presented in this report are consistent with those previously provided in the *URET Delivery 2.1 Conflict Prediction Accuracy Preliminary Report*.

Reported missed alerts for both standard and expanded separation standards have a very low probability (0.0016 for Analysis A; 0.0092 for Analysis B). In the analysis of the nine simulation runs based on standard separation (i.e., Analysis A), only one missed alert was found. That is, two aircraft approached to within five nautical miles of each other horizontally and had less than 2000 feet vertical separation, and URET failed to predict a conflict. The strategic missed alert probability is higher at 0.0408 for Analysis A, and 0.0696 for Analysis B. The false alerts are much more common and possibly a reason for the low missed alert probability, since these two fundamental errors have an inversely proportional relationship (i.e., a false alert probability that is high by definition will provide a low missed alert probability). A conflict probe attempts to trade off between these two errors so that both are acceptable (i.e., such that the occurrence of missed alerts is essentially zero). For Analysis A, the false alert probabilities were predominate at minimum horizontal distances between 5 and 10 nautical miles, with a grand average false alert probability for both red and yellow alerts of 0.8481. When the actual conflict definition was expanded to 10 nautical miles in Analysis B, the false alert probabilities were reduced by approximately one fourth to 0.6451. The false alert probability can also be conditioned by specific alert level (i.e., red or yellow). When the probabilities are conditioned in this manner for Analysis A, the conditional probability for a red alert being false is 0.72, and 0.93 for yellow. In other words, given a red alert, the probability that the alert is false is 0.72 for Analysis A. Similarly, for Analysis B, the conditional probability for a red alert being false is 0.51, and 0.73 for yellow. Finally, the grand average warning time is estimated at 15 minutes and the average conflict start time delta is 222 seconds (3.7 minutes) for Analysis A. There is a small reduction in these times for Analysis B (grand average warning time is estimated at 13.5 minutes and the average conflict start time delta is 130 seconds (2 minutes and 10 seconds), which is to be expected since an aircraft pair will be in violation of ten nautical mile separation earlier than for five nautical mile separation.

The metrics described in this report were designed to be generic so they can be applied to any conflict probe tool. This study developed the metrics, and then focused on their estimation for URET D3. As stated, this included analysis of nine simulations of ZID ARTCC, totaling roughly 4500 aircraft. The method of analysis was specifically designed to allow the reader to decide on how these estimates should be used. This is exemplified by the twofold analysis, designated as Analysis A and Analysis B. The reader may choose estimates in either analysis, or gain insight in the sometimes subtle differences between the two. Furthermore, if the reader wishes to calculate other statistics based on the error events present in the simulations, two extensive appendices are supplied that contain all the partitioned individual counts from all nine simulations. Additional information, not contained in these appendices, can be obtained from the large set of Oracle databases (containing both raw and processed data) resident in the TFM laboratory at the WJHTC upon formal request.

The information provided in this report is valuable to decision makers charged with determining if the URET prototype should be installed in additional ARTCCs, as well as those making an investment decision for a production conflict probe. It should also prove useful to both the developer of the URET prototype and the conflict probe production contractor. Future studies should be conducted, using actual field data adjusted to include predefined conflict situations, to expand the conflict prediction accuracy estimation to include the complete set of generic metrics (e.g., trajectory accuracy and conflict prediction stability) defined in *Generic Metrics and Statistics to Estimate the Conflict Prediction Accuracy of Conflict Probe Tools* and *Plan for Evaluation of the Conflict Probe Programs.*

References

Generic Metrics and Statistics to Estimate the Conflict Prediction Accuracy of Conflict Probe Tools (draft), WJHTC (ACT-250), August 1997

ICP/Enhanced Display System Infrastructure (EDI) System/Segment Specification, Volume I, Part 2 (Work In Progress), Lockheed Martin Air Traffic Management, February 27, 1998

Initial Assessment of Requirements for Mid-Term Conflict Probe Performance: Testbed Status Report (draft), MITRE/CAASD, December 1997

Introduction to Statistical Quality Control, Second Edition, Montgomery, Douglas, C., 1991.

Plan for Evaluation of the Conflict Probe Programs (draft), AUA-540/ACT-250, February 4, 1997

Performance Analysis of AERA Algorithms: Phase II, Volume 3, MITRE/CAASD, November 1984.

Pseudo Aircraft System (PAS) User's Manual, LOGICON/Syscon/Syre, November 4, 1997

User Request Evaluation Tool (URET) Algorithm Assessment Report, ACT-250, February 1997

URET Algorithmic Assessment Report: Interfacility Capability (draft), MITRE/CAASD, December 1997

URET Delivery 2.1 Conflict Prediction Accuracy Preliminary Report (preliminary draft), ACT-250, January 1998

URET Delivery 3 System Testing Tools, Scenarios, and Test Results, MITRE/CAASD, December 1997

Acronyms

ACD	Automated Conflict Detection
ACES	Adaptation Controlled Environment Subsystem
AEC	Algorithm Evaluation Capability
APDIA	Automated Problem Detection Inhibited Area
ARTCC	Air Route Traffic Control Center
ATM	Air Traffic Management
CAASD	Center for Advanced Aviation System Development
CAT	Conflict Analysis Tool
DR&A	Data Reduction and Analysis
DSR	Display System Replacement
D3	Delivery 3
FA	False Alert
GEN	Generation
GPOIU	General Purpose Output Interface Unit
HCS	Host Computer System
ICP	Initial Conflict Probe
IFR	Instrument Flight Rules
JRC	Joint Resources Council
LOA	Letter of Agreement
LVA	Late Valid Alert
MA	Missed Alert
MHS	Minimum Horizontal Separation
MTR	Monitor Test and Recording
MVS	Minimum Vertical Separation
NAS	National Airspace System
PAS	Pseudo Aircraft System
SAR	System Analysis Recording
SMA	Strategic Missed Alert
SOP	Standard Operating Procedure
TCP	Track Conflict Probe
TOD	Top of Descent
TFM	Traffic Flow Management
TGF	Target Generation Facility
UFP	Universal Flight Plan
URET	User Request Evaluation Tool
VA	Valid Alert
WJHTC	William J. Hughes Technical Center
ZID	Indianapolis ARTCC
ZME	Memphis ARTCC

Appendix A: Analysis A Data

(in separate file: ucpa_apa.doc)

Appendix B: Analysis B Data

(in separate file: ucpa_apb.doc)

Appendix C: Sharpness Metric

C. Background

The basic errors associated with a conflict probe are referred to as missed alert and false alert. It is not sufficient to report only a missed alert probability without a corresponding false alert probability (and vice versa), since these two fundamental errors are not independent and furthermore are inversely proportional. As a result, it is possible to reduce either one of these probabilities with an increase to the other. The balance between false alert and missed alert probabilities not only must be acceptable, but balance should be robust in terms of the separation between aircraft and other factors. The need to determine the proper trade off between the two error probabilities prompted ACT-250 to research metrics to model the sensitivity of the conflict predictions.

Conflict probe tools are not the only detection systems that require sensitivity measures. For example, in the manufacturing industry Statistical Quality Control (SQC) Charts[4] are used to detect shifts in the manufacturing process. These charts detect a change or shift in the process average to minimize defective product. Similar to the conflict prediction of a conflict probe, there are two errors associated with the detection of a shift in the process mean, referred to as type I and type II. The type I error probability refers to the probability of detecting a shift when a shift did not occur; this is analogous to the false alert probability for a conflict probe. The type II error is the probability of not detecting a shift when the process mean really did make a shift; this is comparable to the missed alert probability. In the design of control charts, quality engineers trade off between these two errors by drawing curves referred to as the operating characteristic function. These sets of the curves are a plot of the probability of not detecting a shift versus an associated actual shift in the process mean. The quality engineer plots these curves for different chart designs (e.g., different sample sizes) to determine which one best fits the particular process. The curve with the steepest relationship between the probability of missing a shift versus the actual shift magnitude would minimize errors associated with the detection system. The operating characteristic function can be applied to the conflict prediction process for a conflict probe. The sharpness[5] metric is an example of such a function for modeling the precision of a conflict probe's conflict prediction.

C.1 Sharpness

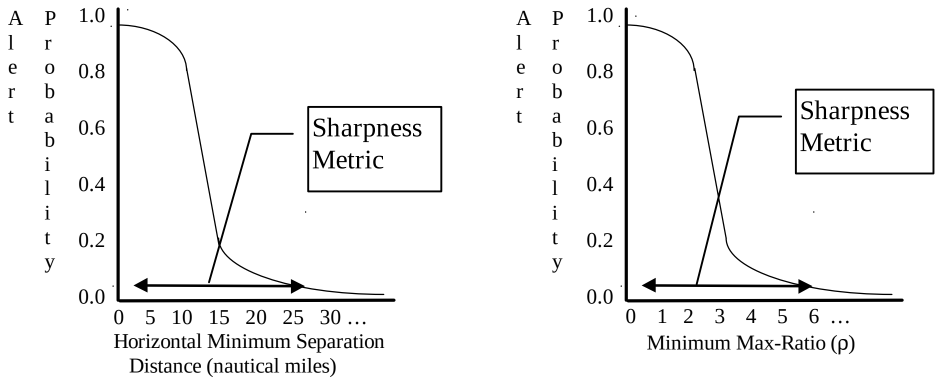
The sharpness metric is a measure of the average sensitivity of a conflict probe's aircraft-to-aircraft conflict predictions. Sharpness measures how quickly the probability of an alert drops, from a value near one to a value near zero, as the aircraft separation increases. To determine sharpness, a performance curve is formed by plotting the probability of a conflict prediction by the conflict probe versus the actual minimum separation distance between aircraft (refer to Figure C.1-1a). The probability of a conflict prediction by the probe is the measure of the likelihood of an alert being presented to the controller for a particular aircraft pair. For our purposes, the actual minimum separation distance is calculated from the HCS's track reports during post-processing. The sharpness metric is calculated by finding the intersection points of a probability close to one and the performance curve, and a probability close to zero and the performance curve. Specifically, the distance along the x-axis between these two points defines the sharpness metric. The precision of the conflict prediction can thus be indicated by the sharpness metric by measuring the steepness of the performance curve. That is, the steeper or more abrupt the incline of the curve, the better the precision of the conflict prediction.

One condition for the "perfect" conflict probe is that the sharpness will equal zero. As shown in Figure C.1-2, the probability curve for the perfect probe will not form a curve at all, but will be a step function. This perfect probe would have a probability of one in detecting a conflict with minimum separation distance of zero up to the specified separation standard. At the separation standard and greater, the perfect probe would have a probability of zero in detecting a conflict. That is, the better the performance of the conflict probe, the smaller the sharpness distance will be.

---

[4] <u>*Introduction to Statistical Quality Control, Second Edition*</u>.

[5] Sharpness expands upon the MITRE/CAASD-defined metric referred to as "crispness" in *Performance Analysis of AERA Algorithms: Phase II, Volume 3*.



(a) Sharpness for aircraft pairs on the same flight levels

(b) Sharpness for all aircraft, captures both vertical and horizontal separation

Figure C.1-1: Example of Alert Probability Versus Minimum Separation Distance

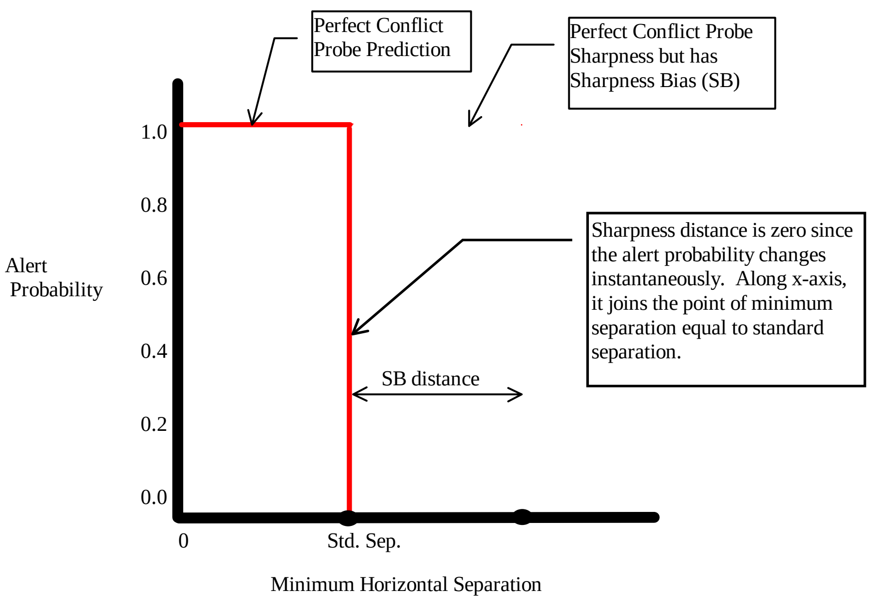


Figure C.1-2: Ideal or Perfect Conflict Probe Performance

Referring to Figure C.1-1a, the distances along the x-axis of the probability curve are the minimum horizontal distances of the aircraft pairs flying on the same flight level only (i.e., with less than standard vertical separation). For this subset of aircraft pairs, the area under the curve is related to the missed and false alert probabilities. The aircraft pairs on the same flight level represent only a subset of the total number of aircraft pairs. In order to consider all aircraft pairs not just on the same flight levels, it is necessary to capture both dimensions of separation on the x-axis (refer to Figure C.1-1b), since the legal separation of aircraft is presented in both the horizontal and vertical dimensions. For the horizontal dimension, the standard separation is given in nautical miles (nominally 5 nautical miles). For the vertical dimension, the standard separation is presented on a much smaller scale (nominally 1000 feet for aircraft at or below flight level 290, and 2000 feet for aircraft above FL290). In other words, an aircraft needs 15 times more separation in the horizontal plane than in the vertical. These two dimensions of separation distances are practically independent, but a conflict takes place only if both are violated simultaneously. A metric has been defined to capture the independent processes in both dimensions into one value that corresponds to the aircraft pair's minimum separation. First, the separation distance in each dimension is normalized, so both values are on the same scale. This is accomplished by dividing the aircraft to aircraft separation by the standard separation for each time synchronized position report. The standard separations may vary depending on the location of the conflict (e.g. 1000 feet below 29000 feet and 2000 feet above). These ratios are expressed in the following set of equations.

The ratio of horizontal separation to standard horizontal separation can be expressed as:

$$l_i = \frac{\left(\sqrt{\left(\left(x_i^a - x_i^b\right)^2 + \left(y_i^a - y_i^b\right)^2\right)}\right)}{d_i} \qquad \textbf{Equation 1}$$

where
d_i = horizontal separation standard for the i th synchronized track data point;
x_i^a = x position of the i th track point of aircraft a in nautical miles;
x_i^b = x position of the i th track point of aircraft b in nautical miles;
and y_i^a , y_i^b are the corresponding y positions

The ratio of vertical separation to standard vertical separation can be expressed as:

$$p_i = \frac{\left|z_i^a - z_i^b\right|}{u_i} \qquad \textbf{Equation 2}$$

where
u_i = vertical separation standard for the i th synchronized track data point;
z_i^a = altitude position of the i th track point of aircraft a in feet;
z_i^b = altitude position of the i th track point of aircraft b in feet.

Next, the maximum value of l and p is calculated for each track point and the minimum from all these maximums is determined for each aircraft pair. The following equation expresses the calculation of the minimum of the maximum ratios.

$$r = \min {}_{i=}^{k}\left[\max\left(l_i,\ p_i\right)\right] \qquad \textbf{Equation 3}$$

where
i = current i th track point ;
k = total number of track points.

The unitless distance, r , referred to as the minimum max-ratio of separation, combines both dimensions of separation and directly corresponds to standard separations. By definition, if r is less than 1, there exists a violation of standard separation, and if r is equal to or greater than 1 there cannot be a violation of standard separation.

Referring to Figure C.1-1b, the probability of an alert is plotted against the minimum max-ratio, resulting in a curve that includes all aircraft pair encounters not just at the same flight level. The associated sharpness distance is expressed as a minimum max-ratio value. Furthermore, the alerts with associated conflicts (referred to as valid alerts) will produce the curve to the left of r equal to 1 in Figure C.1-1b, and alerts without associated conflicts (referred to as false alerts) produce the curve starting at and to the right of r equal to 1.

The separation measure defined as the minimum max-ratio is best illustrated by the following example. An aircraft pair are initially separated by 50 nautical miles or greater in the horizontal, but are flying below flight level 290 at adjacent flight levels so are separated by the vertical separation. From Figure C.1-3, the l value or horizontal separation ratio starts at above 10 and is decreasing. Therefore, the aircraft start on converging routes in the horizontal dimension. The vertical separation ratio, p, begins at one and then decreases to below one before increasing again. Actually what is happening is the aircraft at the higher altitude starts a descent around the point where the aircraft pass horizontally. The vertical separation reduces as the aircraft passes through the other's altitude level and starts increasing as the descending aircraft continues. As the aircraft pass by each other in the horizontal within standard separation, the vertical max-ratio drops below one forming the duration of the conflict encounter. After the conflict ends, the vertical separation is now the maximum and remains the max-ratio for the completion of the flight.

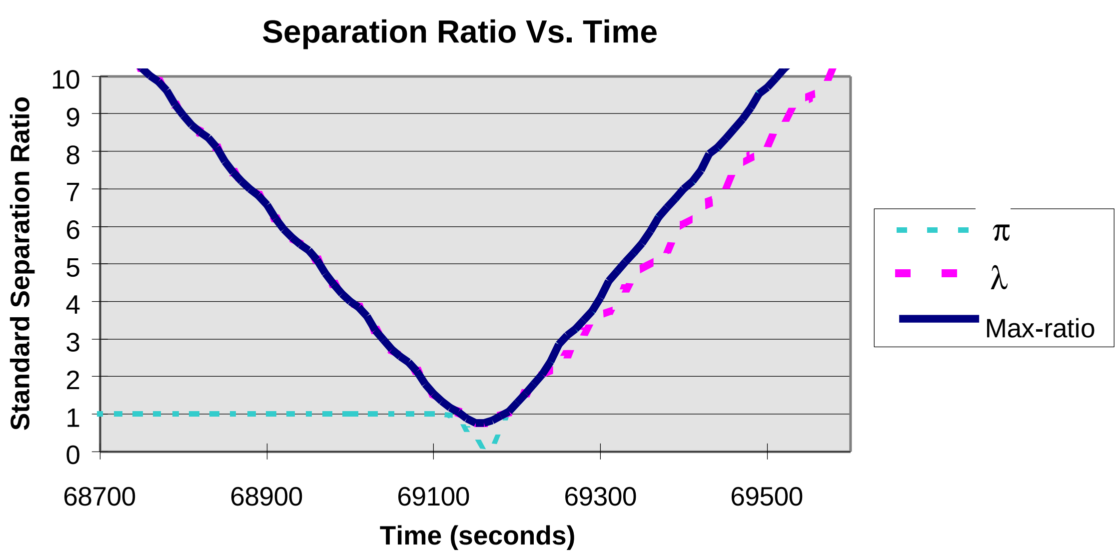


Figure C.1-3: Plot for Max-Ratio Against Time

C.2 Sharpness Bias

In defining the perfect probe, however, a sharpness of zero is not sufficient since the alert probability curve could be offset along the x-axis. This offset or separation distance bias, called sharpness bias, is defined as the minimum max-ratio corresponding to an alert probability of 0.5 minus one. The perfect probe would have an sharpness bias of zero, since the minimum max-ratio at 0.5 probability of alert would be one. This is illustrated in Figure C.1-2 where dotted step function represents a sharpness of zero but has a significant sharpness bias.

The sharpness bias represents the conflict probe's built in tolerances used for conflict prediction. A positive value for sharpness bias would indicate the sharpness curve has a bias greater than the defined separation for conflicts and would favor false alerts over missed alerts. A negative value for sharpness bias indicates the sharpness curve has a bias less than the defined separation for conflicts and would favor missed alerts over false alerts. In other words, a positive value of sharpness bias would indicate that the conflict prediction tends to have more false alerts than missed alerts on average. Once again the extreme case is illustrated in Figure C.1-2 where the dotted step function falls to the right of standard separation (using the minimum max-ratio it would be equal to 1.0). For this example, the probe had a perfectly sharp conflict prediction (i.e. sharpness equal to zero), but it was bias to the right which means false alerts must occur without missed alerts.

C.3 Estimation of Sharpness for URET D3

To calculate sharpness and sharpness bias, the HCS track reports from the nine simulations were compared and the actual horizontal and vertical separations were used to calculate the minimum max-ratio of all aircraft combinations[6]. Next, the URET alerts were matched with the minimum max-ratio values calculated from the track reports. Therefore, the probability of an alert is actually a conditional probability of an alert with a certain range of r (i.e. minimum max-ratio) given the existence of aircraft pairs with a certain range of r. To estimate this conditional probability, the number of alerts presented with a certain range of r (i.e. minimum max-ratio) are calculated from HCS track and are divided by the total number of aircraft pairs with the same range of r.

For this study, an interval of 0.1 r was used for calculating the sharpness curve. To estimate the curve, a histogram is formed with probabilities for each 0.1 interval. To calculate sharpness, the distance along the x-axis is calculated between two r values chosen from translating the points from a probability close to 1 to a probability close to 0. The parameters chosen for this study were 0.99 and 0.10. These values were chosen to capture the distance sharpness is designed to measure, namely the sensitivity of conflict predictions to the true separation of aircraft. The upper threshold is 0.01 less than 1.0 and lower threshold of 0.10 is ten times that probability distance from 0 probability. There are several reasons for the difference in the thresholds. One reason is to emphasize the greater significance of a missed alert compared to a false alert. Another reason is to minimize the large variations on sharpness caused by the greater distances present at the upper tail of the probability curve (i.e. above 0.10). Interpolation is used to translate the probability thresholds to the appropriate r value on the x-axis (details are discussed in *Generic Metrics and Statistics to Estimate the Conflict Prediction Accuracy of Conflict Probe Tools*). For example, in Figure C.3-1 two conditional alert probability versus minimum max-ratio plots are presented for simulation number four. The first set of curves in light gray represent the alerts for Analysis A where the conflicts were based on standard separation distances. The next set of curves in dark red represent the alerts for Analysis B where the conflicts were based on expanded separation distances (i.e. 10 nautical miles horizontal separation). The dotted curves are the plot of the actual probability estimates for each interval. Since the actual probability estimates are hard to compare because of the sampling noise, a set of smoothed moving average plots are also presented in Figure C.3-1. As expected, the sharpness for Analysis B is significantly smaller than Analysis A in this simulation run. The difference in sharpness is around 1.5 r. The sharpness bias is also smaller for Analysis B, since the alert probability curve that crosses the 0.5 probability is around 1 r value smaller than Analysis A.

The sharpness and sharpness bias metrics were estimated with 95% confidence intervals for all nine simulation runs for both Analysis A and B (see Table C.3-1). The average sharpness for Analysis A and B was around 3 and 1.4, respectively. The average sharpness bias for Analysis A and B was around 1.1 and 0.2, respectively. As expected, URET predicts conflicts with a significantly smaller sharpness when considering a conflict around twice the standard horizontal separation (i.e. 10 nautical miles and standard vertical). Therefore, when considering 10 nautical mile conflicts, the conflict predictions are much more

[6] Actually the minimum max-ratio is calculated for all the aircraft combinations that had some time overlap and passed the TCP gross filter (refer to section 2.3.1).

precise compared to the standard 5 nautical mile conflicts. The sharpness bias suggests that the URET conflict predictions only slightly favor false alerts for Analysis B. In other words, URET is predicting the expanded 10 nautical mile conflicts with significantly more precision and a very little bias compared to conflict predictions of standard separation conflicts.

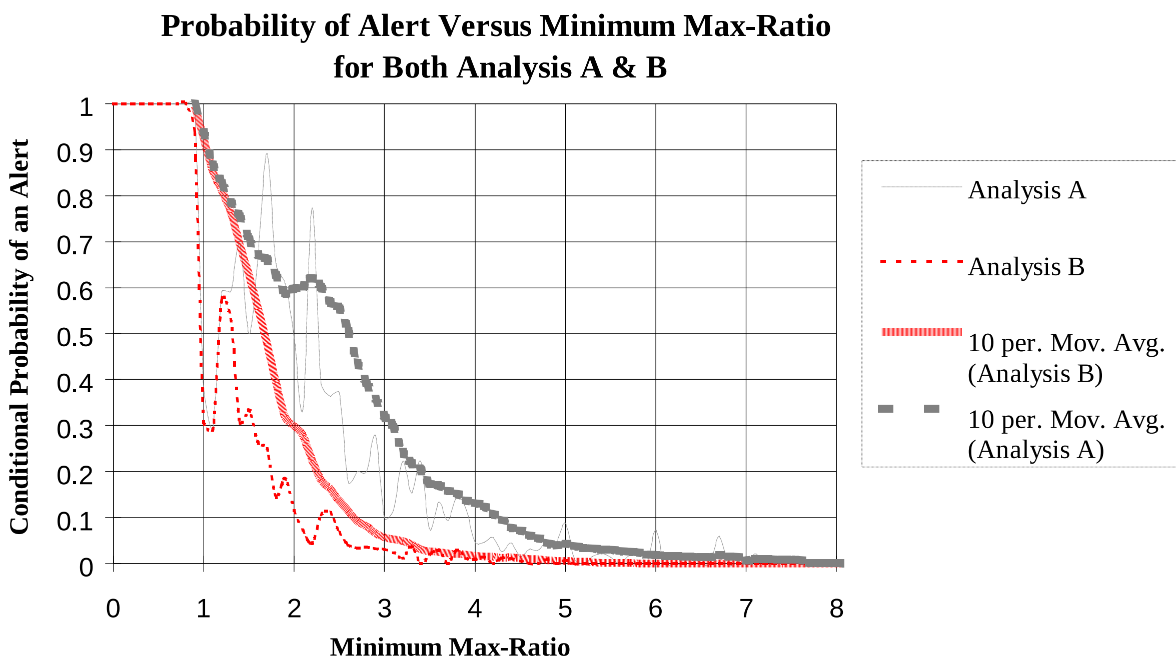


Figure C.3-1: Plot of Conditional Probability of an Alert Versus Minimum Max-Ratio

Simulation #	S - Analysis A	SB - Analysis A	S - Analysis B	SB - Analysis B
1	3.0279	0.9747	1.5817	0.308
2	2.8935	0.9758	1.344	-0.0047
3	2.9873	1.2775	1.2338	0.3416
4	3.0268	1.2718	1.5731	0.3136
5	3.471	1.2654	1.3362	0.2377
6	2.6722	0.9742	1.4069	0.2145
7	2.3913	0.9227	1.0341	-0.0154
8	3.4426	1.3219	1.5982	0.2713
9	2.9957	0.9734	1.154	-0.0185
Average	2.9898	1.1064	1.3624	0.1831
Std. Deviation	0.3369	0.1702	0.1996	0.152
Upper Limit*	3.2487	1.2372	1.5159	0.3
Lower Limit*	2.7309	0.9756	1.209	0.0663

* Based on a 95% confidence interval using a Student t distribution.

Table C.3-1: Sharpness (S) and Sharpness Bias (SB) Results

C.4 Summary

This appendix is included to provide information on the definition and estimation of the sharpness metric. The sharpness metrics are designed as an aggregate metric to measure the precision and sensitivity of the conflict predictions. These metrics are recommended as a relative measure when comparing the performance of different conflict probes, different site adaptations, or different parameters used in a conflict probe, and thus can be useful in the development, evaluation and specification of a conflict probe.

There are many potential applications of the sharpness metrics. For example, different sites with various field adaptations and traffic mixes will certainly create different demands on the performance of a conflict probe. The sharpness and sharpness bias could provide a sensitivity measure on how the performance varies from site to site. It could also provide a useful gauge to the developer on what direction to adjust conflict probe parameters to achieve a level performance. The most effective use of this metric requires further detailed designed experiments to be conducted to determine what factors have a statistically significant effect on sharpness and thus the conflict prediction precision.